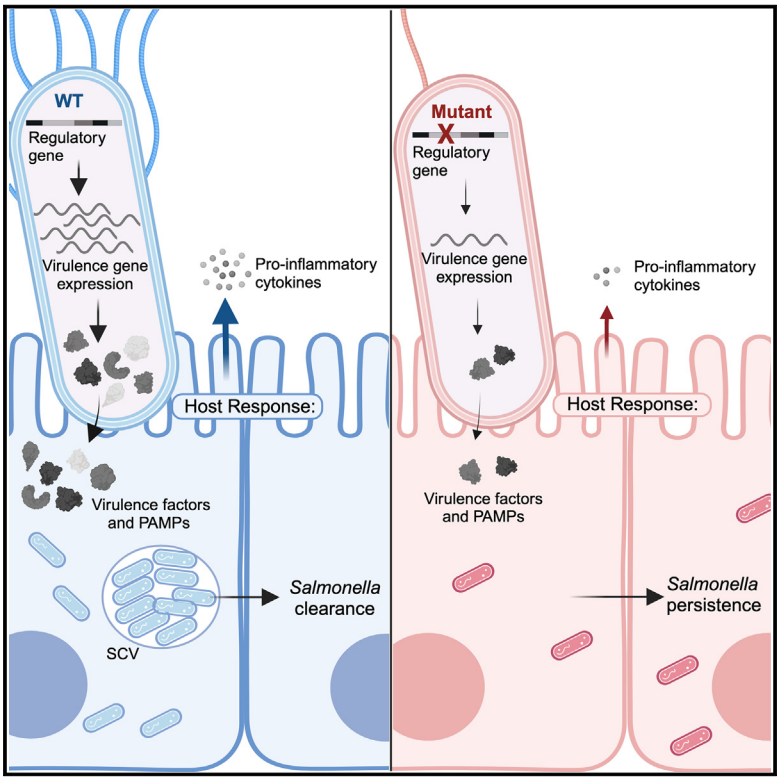
Article

Cell Host & Microbe

Persistent *Salmonella* infections in humans are associated with mutations in the BarA/SirA regulatory pathway

Graphical abstract



Authors

Alexandra Grote, Bar Piscon, Abigail L. Manson, ..., Jonathan Livny, Ashlee M. Earl, Ohad Gal-Mor

Correspondence

livny@broadinstitute.org (J.L.), aearl@broadinstitute.org (A.M.E.), Ohad.Gal-Mor@sheba.health.gov.il (O.G.-M.)

In brief

Grote et al. show that adaptations in global regulatory genes arise during persistent *Salmonella* infections in humans. These adaptations lead to decreased virulence gene expression, attenuated pathogenicity, and a weaker immune response in an acute salmonellosis model but to long-term colonization and shedding in a persistent mouse model.

Highlights

- *Salmonella* global regulators are frequently mutated during persistent human infection
- The BarA/SirA virulence regulatory pathway was most frequently mutated
- *barA/sirA* mutants were less virulent and elicited a weakened host immune response
- *barA/sirA* mutants colonized mice and were shed during persistent salmonellosis

Grote et al., 2024, Cell Host & Microbe *32*, 79–92
January 10, 2024
https://doi.org/10.1016/j.chom.2023.12.001

CellPress

Cell Host & Microbe

CellPress

Article

Persistent *Salmonella* infections in humans are associated with mutations in the BarA/SirA regulatory pathway

Alexandra Grote,[1] Bar Piscon,[2,3] Abigail L. Manson,[1] Boaz Adani,[2] Helit Cohen,[2] Jonathan Livny,[1,*] Ashlee M. Earl,[1,*] and Ohad Gal-Mor[2,3,4,*]

[1]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[2]Infectious Diseases Research Laboratory, Sheba Medical Center, Tel-Hashomer, Israel
[3]Department of Clinical Microbiology and Immunology, Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
[4]Lead Contact
*Correspondence: livny@broadinstitute.org (J.L.), aearl@broadinstitute.org (A.M.E.), Ohad.Gal-Mor@sheba.health.gov.il (O.G.-M.)
https://doi.org/10.1016/j.chom.2023.12.001

SUMMARY

Several bacterial pathogens, including *Salmonella enterica*, can cause persistent infections in humans by mechanisms that are poorly understood. By comparing genomes of isolates longitudinally collected from 256 prolonged salmonellosis patients, we identified repeated mutations in global regulators, including the *barA/sirA* two-component regulatory system, across multiple patients and *Salmonella* serovars. Comparative RNA-seq analysis revealed that distinct mutations in *barA/sirA* led to diminished expression of *Salmonella* pathogenicity islands 1 and 4 genes, which are required for *Salmonella* invasion and enteritis. Moreover, *barA/sirA* mutants were attenuated in an acute salmonellosis mouse model and induced weaker transcription of host immune responses. In contrast, in a persistent infection mouse model, these mutants exhibited long-term colonization and prolonged shedding. Taken together, these findings suggest that selection of mutations in global virulence regulators facilitates persistent *Salmonella* infection in humans, by attenuating *Salmonella* virulence and inducing a weaker host inflammatory response.

INTRODUCTION

Different clinically important bacterial pathogens, including *Mycobacterium tuberculosis*, *Helicobacter pylori*, *Pseudomonas aeruginosa*, and *Salmonella enterica*, are able to establish persistent infections in humans. These persistent bacteria can evade the host immune system and killing by antibiotics and are associated with the emergence of new antibiotic resistant strains.[1] *S. enterica* is a Gram-negative, facultative intracellular pathogen with over 2,600 antigenically distinct serovars,[2] the majority of which are considered non-typhoidal serovars (NTSs) and frequently colonize food-producing and wild animals. Human infection by *S. enterica* serovars may cause different clinical outcomes, including asymptomatic colonization, gastroenteritis in the terminal ileum and colon, invasive systemic disease and bacteremia, and enteric (typhoid) fever. The precise outcome of a *Salmonella* infection is dependent on the characteristics of the infecting serovar and the immunological status of the infected individual.[3] *S. enterica* infections still pose a significant global clinical challenge with over 27 million cases of enteric fever[4] and 78.7 million cases of gastroenteritis annually.[5]

The two main hallmarks of *S. enterica* pathogenicity are the ability to invade nonphagocytic cells and to replicate within phagocytic and nonphagocytic host cells. These phenotypes are dependent on the function of two separate type 3 secretion systems (T3SSs) and their designated secreted effector proteins. T3SS-1, which is encoded on *Salmonella* pathogenicity island (SPI)-1, is expressed when *Salmonella* are extracellular and mediates *Salmonella*'s invasion of nonphagocytic cells.[6,7] Additionally, the products of SPI-1 genes are responsible for regulation of the host immune response,[8,9] and induction of neutrophil recruitment during enteric colitis, which lead to reduction and alteration of the intestinal microbiota.[10,11] In contrast, T3SS-2, encoded on SPI-2, is induced when *Salmonella* is intracellular and is required for intracellular survival and replication.[12,13]

SPI-1 gene regulation is complex and involves multiple regulators, encoded both within and outside of SPI-1,[14] that control gene expression at the transcriptional and post-transcriptional levels.[14] The BarA/SirA two-component regulatory system (TCRS) consists of BarA, the sensor kinase, and SirA, the response regulator.[15] The BarA/SirA system can activate the SPI-1 transcriptional regulators HilD and InvF.[16,17] In addition, the BarA/SirA system promotes SPI-1 gene expression by activating the expression of *csrB/C*, two non-coding regulatory RNAs that counteract CsrA-mediated repression of *hilD* transcript.[15,18] This BarA/SirA TCRS is well conserved across different bacteria and has orthologs in *Pseudomonas* species (GagA/GacS), *Vibrio cholerae* (VarA/VarS), *Erwinia*

carotovora (ExpA/ExpS), *Legionella pneumophila* (LetA/LetS), and *Escherichia coli* (UvrY/BarA) where it is similarly involved in regulation of virulence, antibiotic production, motility, and biofilm production.[19]

Although it is well known that typhoidal serovars cause persistent salmonellosis,[3,20,21] we previously showed that NTS can also establish symptomatic or asymptomatic prolonged infections in humans at a frequency of at least 2.2%, which can last months to years.[22] Although the underlying mechanisms of *Salmonella* persistence are poorly understood, studies of NTS infections in humans and mouse models have shed some light on host, bacterial, and environmental factors that contribute to persistent salmonellosis, including gender and age,[23–26] and antibiotic treatment.[27] In our previous study of paired whole-genome sequences of early and late isolates from 11 patients with *S.* Typhimurium infections, single-nucleotide polymorphisms (SNPs) were found in global virulence regulatory genes, including *dksA*, *rpoS*, *hilD*, *melR*, *rfc*, and *barA*.[22] Nonetheless, no convergence at the SNP or gene level was identified between patients. Similarly, in a study of whole genomes from an *S.* Typhimurium infection lasting 55 days, a nonsynonymous mutation was found in *flhC*, encoding a master regulator of flagellar biosynthesis.[28] These singular examples point to a role for key virulence regulators in the development of persistent salmonellosis, but it is still unclear what molecular pathways enable persistence or how this process evolves during human infections.

RESULTS

Persistent infections are caused by a diverse set of *Salmonella* serovars

To profile adaptations leading to persistence across NTS, we collected and whole-genome sequenced all *S. enterica* isolates obtained from patients in Israel presenting with persistent symptomatic salmonellosis between 1995 and 2012, drawn from a large retrospective study of *Salmonella* persistence, including 48,345 culture-confirmed salmonellosis cases (NCBI BioProject ID: PRJNA847966)[22] (Table S1). Each of the 256 patients was represented by 2–5 independent isolates obtained at different times during their infection that lasted between 30 and 2,001 days from the first culture-confirmed diagnosis. Of the 639 isolates sequenced as part of this study, analysis using Centrifuge[29] revealed that data from eight samples (from three different patients) showed high amounts of non-*Salmonella* bacterial DNA, likely due to co-infection or sample contamination, and were therefore excluded from further analysis. Sequence reads from each of the remaining 631 isolates were aligned to the reference genome of *S.* Typhimurium SL1344 (NC_016810.1) to identify variants, including SNPs, which were used as input into a maximum likelihood-based phylogenetic analysis to solve the evolutionary relationships among isolates. Mapping the results of laboratory serotyping onto the phylogeny revealed largely congruent results, with each of the 49 predicted serovars corresponding to a closely related cluster on the phylogenetic tree (Figure 1A).

These analyses showed that persistent salmonellosis can be caused by many NTS serovars. However, in support of our previous work,[22] we found that serovars Mbandaka, Bredeney, Infantis, and Virchow were particularly common among persistent isolates, indicating their apparent propensity to cause persistent infections.

Most recurrent infections are due to persistence rather than reinfection

Although most same-patient isolates fell within the same serovar, we sought to differentiate cases of persistent infection by the same strain from reinfection by different strains of the same serovar. To do this, we performed pairwise comparisons of isolates from the same patient obtained at different time points using variants called against the *S.* Typhimurium SL1344 reference genome. The first culture-confirmed isolate was considered the "early" isolate, and the subsequent isolates were referred to as the "late" isolates. 31 patients had isolates that were separated by more than 200 SNPs, likely representing reinfections or original non-clonal infections, rather than persistence by a single infecting clone. The remaining isolates from 225 different patients were separated by 36 SNPs or fewer, an expected range based on previously published SNP rates (1–10 SNPs per genome per year) and the maximum length of infection in our study (Figure 1B).[22] This threshold included 72 patients (28%) with zero SNPs separating their isolates. The number of SNPs accumulating over time, on average, across pairs was found to be 9.58 substitutions per year, similar to the number expected based on SNP rates calculated for *S.* Typhimurium.[22,28]

Persistent isolates show stable plasmid and AMR gene carriage within a patient during persistence

Plasmids are a known, flexible component of the *Salmonella* genome that facilitate horizontal gene transfer and often carry antibiotic resistance genes, which could potentially contribute to long-term persistence in an infected host.[3,23,30,31] To determine the extent to which plasmid or antibiotic resistance gene (ARG) content was associated with persistence, we first characterized plasmids and ARG content in each isolate using MOB-suite[32] and Resistance Gene Identifier (RGI),[33] respectively. Overall, MOB-suite predicted 1,193 plasmids among the 639 *Salmonella* isolates (1.87 plasmids per isolate, on average), which were clustered into 96 distinct homology groups based on genomic distance estimation using Mash (Table S2, Sheet 1).[34] The most prevalent plasmid groups tended to be serovar specific and widely held among different isolates of that serovar, supporting prior work showing a serovar-specific distribution of plasmids among Salmonellae, including the pSLT virulence plasmid of *S.* Typhimurium, the pSEN plasmid of *S.* Enteritidis,[35] and the pESI plasmid of *S.* Infantis (Figure S1B).[36,37] Additionally, we identified a previously unreported association between the p2.3 plasmid and *S.* Enteritidis in the majority of persistent isolates of this serovar. Conversely, plasmids pSH14-028_2 and pSAN1-1677 (in red and orange, respectively) did not appear to be serovar specific, but rather were found in 133 and 93 isolates, respectively, across many diverse serovars of *Salmonella* in the dataset of persistent isolates.

Interestingly, when we compared plasmid carriage across early and late same-patient isolates using plasmid group assignments (Table S2, Sheet 2), we found plasmid content to be generally stable over time, with only little evidence for plasmid gains and losses during persistent human infection (Figure S1A). Moreover, among same-patient isolates changing in plasmid content over time, there was no association between persistence and the number of gains

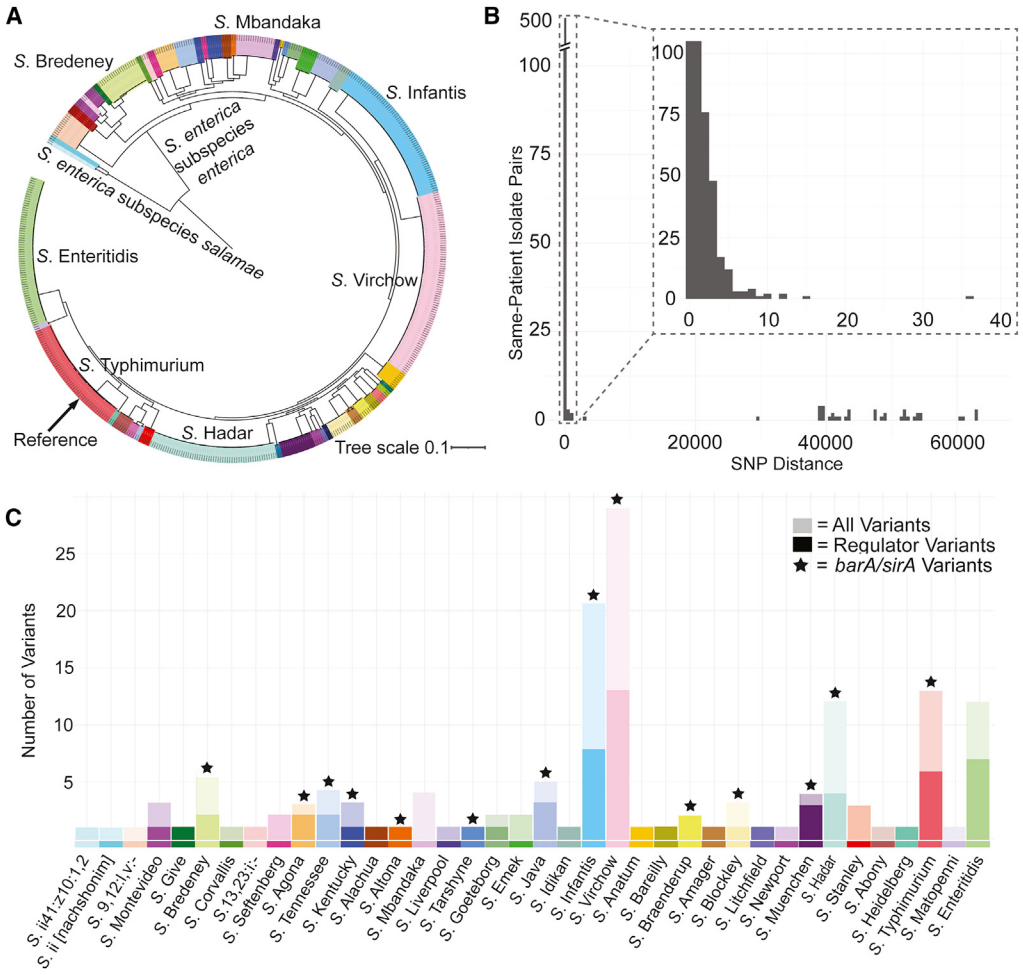


Figure 1. Genetic diversity of *S. enterica* isolates responsible for persistent infections

(A) Phylogeny of 562 *S. enterica* isolate genomes from human persistent infections. The colored ring represents serovar information as determined by serotyping according to the White-Kauffmann-Le Minor scheme.[2]

(B) Distribution of SNP distances between 545 isolate pairs obtained from the same patient at different time points.

(C) Variants, including SNPs, insertions, and deletions, in all genes, and in genes with the GO term for "Regulation of Biological Process" (GO:0050789), between early and late same-patient isolates. Bars are colored by serovar, with light tone representing all variants and dark tone representing variants in genes with GO:0050789. The star indicates the serovars presenting variants in *barA* and/or *sirA*.

or losses of a plasmid group, or plasmid group frequency (Supplemental Data, Figures S1C and S1D). Similarly, we found that ARG content of most same-patient isolates (whether predicted to be plasmid or chromosomally encoded), remained relatively stable. Using RGI,[33] we first identified antibiotic resistance genes in each isolate using a read-based approach. We then assessed the frequency of predicted antimicrobial resistance (AMR) genes in both early and late isolates and found them to be highly similar, with an average of 42.2 and 42.0 AMR resistance genes in early and late isolates, respectively (Figures S1E and S1F). We focused on a smaller group of AMR genes, known to be involved in antibiotic resistance in *Salmonella* (Table S3, Sheet 2), and assessed the relative frequency of those AMR genes between early and late isolates and, again, found highly similar frequencies between early and late isolates, with an average of 7.23 and 7.11 AMR genes, respectively. We also assessed whether resistance gene carriage increased over the duration of infection in late isolates and found that AMR genes did not appear to increase (or decrease) with time of persistence (Figures S1G and S1H). Thus, we concluded that AMR gene carriage, similar to plasmid repertoire, appears to be relatively stable within patients during persistent infection, even over long periods of time (~200 days).

Persistent *Salmonella* isolates are enriched for genetic changes in global regulators

We hypothesized that some *Salmonella* adaptations enabling long-term infection in human hosts would be conserved across serovars and patients at the gene and/or functional level. To search for such adaptations, we annotated and compared the locations of intergenic or coding variants, including SNPs, insertions, and deletions, which differed between early and late same-patient isolates across all cases. Despite 388 variable sites in coding regions, only a small fraction (5.5% or 259 of 4,634) of genes were found to contain genetic variation, between

Table 1. Table of most commonly mutated genes

Gene name	Annotation	Regulator?	Count
barA	signal transduction histidine-protein kinase BarA	yes	16
sirA	response regulator SirA	yes	11
HAD6792910	DUF3300 domain-containing protein	no	6
rpoS	RNA polymerase sigma factor RpoS	yes	5
ramR	regulation of MDR efflux	yes	5
shdA	host colonization factor	no	4
gyrA	DNA gyrase subunit A	no	4
iolR	myo-inositol utilization transcriptional regulator IolR	yes	3
melB	melibiose carrier protein	no	3
rpoC	DNA-directed RNA polymerase subunit beta	no	3
nifJ	pyruvate-flavodoxin oxidoreductase	no	3
dksA	RNA polymerase-binding transcription factor DksA	yes	3
yjiE	HTH-type transcriptional regulator YjiE	yes	2
hfq	RNA-binding protein Hfq	yes	2
dcuA	anaerobic C4-dicarboxylate transporter DcuA	no	2
iamb	maltoporin	no	2
hupA	DNA-binding protein HU-alpha	no	2
tcp	methyl-accepting chemotaxis citrate transducer	no	2
malT	HTH-type transcriptional regulator MalT	yes	2
rcnA	nickel/cobalt efflux system RcnA	no	2
alaS	alanine-tRNA ligase	yes	2
csiD	protein CsiD	no	2
proQ	RNA chaperone ProQ	yes	2
HAD6507795	TetR family transcriptional regulator	yes	2
scsB	DUF255 domain-containing protein	no	2
speFL	leader peptide SpeFL	no	2

Genes with variants between early and late same-patient isolates identified in at least two different patients. Annotation, gene name, whether that gene is a regulatory gene, and the number of different patients with a variant in that gene are listed.

early and late same-patint isolates. Of these genes, 26 were found to contain variants between the early and late isolates in at least two different patients, including both synonymous and nonsynonymous mutations (Table 1; Figure 2A). There were an additional 135 variable intergenic sites; however, only five intergenic regions had variants in two or more patients and the only region with variants in more than two patients was downstream of the gene encoding a YjiH-Family protein, HAD6495080, which was shared by four patients (Table S1).

To determine convergence at a pathway level, we performed a GO term enrichment analysis of genes acquiring nonsynonymous or nonsense mutations and found that all 22 enriched (false discovery rate [FDR] less than 0.05) Biological Process GO terms were involved in regulation and transcription, indicating a significant trend toward variation in regulatory processes during persistence (Table S4, Sheet 1). To explore this further, we interrogated the GO term 0050789, regulation of biological process, which was enriched (FDR of 0.0140) in nonsynonymous variants arising in isolates from diverse serovars. Nearly a quarter (103 of 423) of all observed variants were in genes with this GO term, and nearly half (68 of 150) of patients with at least one variant had a variant in a gene with this GO term (Figure 1C; Table S4, Sheet 2). Of the 26 genes listed in Table 1, 11 were assigned GO term GO:0050789 although we noted that additional genes in Table 1 have also been implicated in regulation in *Salmonella*, including *malT*. Therefore, we concluded that genetic variation emerging over time in these global regulatory genes was common across many diverse *Salmonella* serovars (Figure 1C). We also observed a number of genes implicated in antibiotic resistance showing repeated variation among patients, including *gyrA*, *tetR*, and *ramR*. An SNP in the *gyrA* gene was acquired in the late isolate during persistence in four different patients. In patient 260, the SNP is located within the quinolone resistance-determining region, causing a glycine to aspartate substitution at position 81, which resulted in nalidixic acid resistance in the late isolate (data not shown).

Although variants between same-patient early and late isolates were found at numerous genomic loci, variants in two genes, *barA* and *sirA*, were overwhelmingly the most frequent (Table 1; Figure 2A). Strikingly, we observed an accumulation of 27 variants in *barA* and *sirA* in 24 different patients, across 14 different *S. enterica* serovars (Table 1). *Salmonella* strains in two patients gained mutations in both *barA* and *sirA*, and one patient accumulated two different *barA* mutations. In addition, although found in only a single patient, one pair of isolates had a mutation in *hilD*, one of the transcriptional regulators of SPI-1, activated by the BarA/SirA system.[15] All mutations

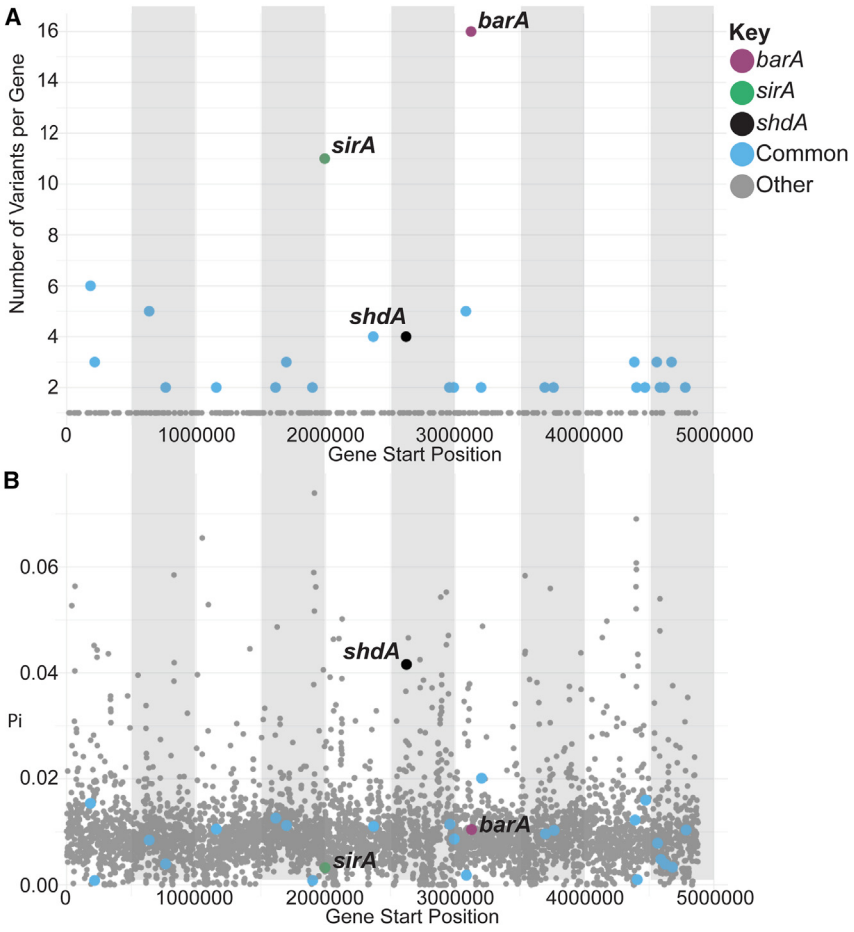


Figure 2. Most commonly mutated genes do not have high nucleotide diversity

(A) The number of distinct variants between early and late same-patient isolates per gene for all genes that contain variants, plotted by gene start position. (B) Average nucleotide diversity, π, per gene in the reference genome, plotted by gene start position. *barA* in purple, *sirA* in green, *shdA* in black, common genes (genes with variants in at least two patients) in light blue, and all other genes in gray.

accumulating over time in these genes were predicted to lead to either missense or nonsense changes.

Convergent evolution in BarA/SirA and other regulatory genes is not due to overall increased mutation rate at these loci

Although the higher frequency of sequence variation between early and late isolates in the genes listed in Table 1 suggested a selection for these variations during persistence, we reasoned that it may also be due to a higher overall mutation rate in these loci. To determine if this was the case, we compared the nucleotide diversity, π,[38] for *barA* and *sirA*, as well as the other genes in Table 1, to that of all other genes in the *S.* Typhimurium SL1344 reference genome using (1) all possible isolate pairs within our collected set of early *Salmonella* isolates and (2) a more diverse set of 875 *S. enterica* complete genomes from RefSeq that were clustered at 99% kmer similarity down to a set of 174 representative genomes (Table S5, Sheet 1). In order to generate an alignment for each gene, we aligned reciprocal best BLAST hits between each RefSeq genome and our reference. For each gene's alignment, we calculated π (Figure S2). The average nucleotide diversity of genes calculated from either set of genomes pointed to the same conclusion. All but two genes in Table 1 were within one standard deviation of the mean nucleotide diversity of all genes in the genome (Figure 2B): *shdA*, (π = 0.042) encoding an outer membrane fibronectin-binding protein previously shown to be highly diverse across *Salmonella*,[39] and *rcnA* (π = 0.020), a gene encoding a nickel/cobalt efflux system (Table S5, Sheet 2). Taken together, these results suggest that variations in *barA* and *sirA*, as well as the majority of other genes in Table 1, reflect convergent evolution of diverse serovars in multiple patients.

BarA and SirA variants in persistent isolates lead to downregulation of virulence genes encoded in SPIs 1 and 4

Because BarA and SirA are known virulence regulators, we next sought to characterize the transcriptional consequences of sequence variations in the *barA* and *sirA* genes. We selected four pairs of early and late same-patient isolates separated by the fewest number of variants, including either a nonsynonymous SNP or deletion in *barA* (2 pairs) or *sirA* (2 pairs) (Figure 3A), and compared their transcriptomes by RNA-seq, following growth to the mid-exponential phase in SPI-1-inducing conditions widely used to mimic the intestinal milieu.[40–45] Although these pairs of isolates were separated by only 1–2 mutations (including the *barA* or *sirA* mutation), we found dozens of genes differentially expressed between each pair of same-patient isolates. Strikingly, the sets of downregulated genes were largely overlapping, with numerous genes encoded in SPI-1 and SPI-4 and genes encoding their associated effector proteins significantly downregulated in all four patients' isolates (Figures 3B and S3; Table S6). The only gene upregulated in late isolates from all four patients pairs was *ompF*, which encodes the OmpF porin protein. Of note, the two isolate pairs with variations in *sirA* shared ten upregulated genes, not observed in the *barA* mutant pairs, including four involved in propionate degradation, *prpBCDE*. It has recently been shown that propionate metabolism is linked to intestinal expansion by *S.* Typhimurium in the inflamed gut.[46] Taken together, these findings suggest that independent mutations in *barA* and *sirA* acquired by late isolates in independent patients lead to largely (although not completely) overlapping transcriptomic changes, which include the downregulation of numerous SPI-1- and SPI-4-encoded factors and associated effectors.

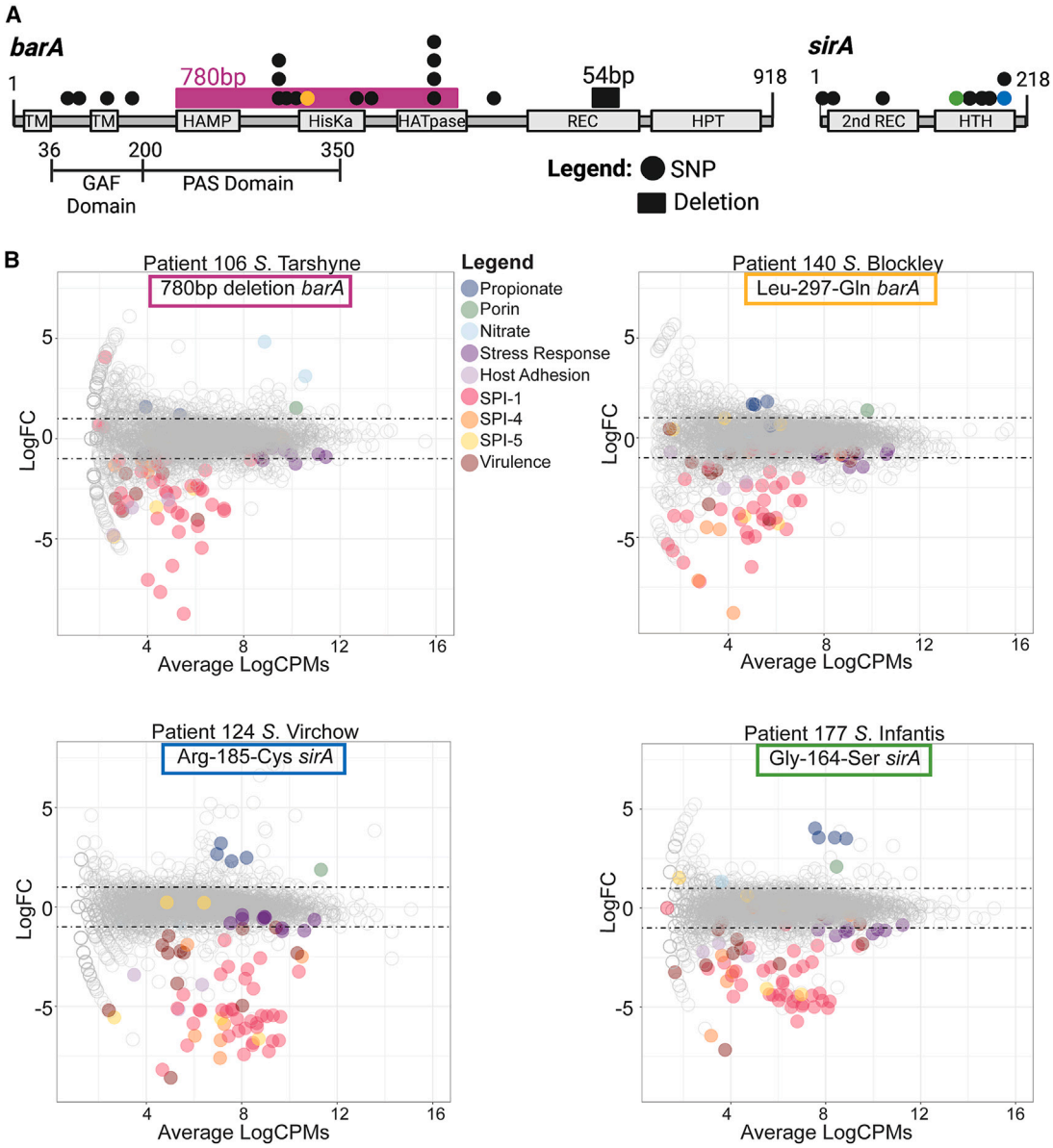


Figure 3. Differentially expressed *Salmonella* genes in patients with *barA*/*sirA* variants

(A) Diagram of the BarA and SirA proteins with domains indicated by boxes: TM, transmembrane domain; HAMP, histidine kinase, adenyl cyclase, methyl-binding protein, phosphatase domain; HisKa, dimerization and phosphoacceptor domain; HATpase, histidine kinase-like ATPase; REC, primary receiver domain; HPT, secondary transmitter domain; 2nd REC, secondary receiver domain; HTH, helix-turn-helix domain. SNPs are indicated by circles and deletions by rectangles, the four colors indicate the SNPs and deletion that were tested in the mouse model.

(B) MA plots showing log fold change (logFC) vs. log average expression (logCPM) for four different patients: two with *barA* mutations (patients 106 and 140) and two with *sirA* mutations (patients 124 and 177). Genes of interest that were found to be differentially expressed in either both *barA* variant strains, both *sirA* variant strains or in all four variant strains are colored based on annotation: propionate degradation (blue), porin (green), nitrate reduction (light blue), stress response (purple), host adhesion (light purple), SPI-1 (red), SPI-4 (orange), and SPI-5 (yellow), virulence (maroon). The colored box around the location matches the color of the SNP or deletion in (A). The complete list of DEGs is shown in Table S6.

Acquired mutations in barA and sirA attenuate *Salmonella* virulence in the acute salmonellosis mouse model

Because genes encoded in SPI-1 and SPI-4 are required for *Salmonella* virulence, we hypothesized that the sequence variations in *barA* and *sirA* would lead to a decrease in *Salmonella* pathogenicity. To test this, we performed competitive infection in streptomycin pre-treated C57BL/6 mice, an established acute salmonellosis model.[47] For each of the four pairs of isolates for which we performed RNA-seq above, equal amounts of the differentially marked early and late same-patient isolates were used to co-infect mice. At 4 days post infection (p.i.), we plated spleen, liver, cecum, and colon homogenates onto antibiotic selective plates and calculated

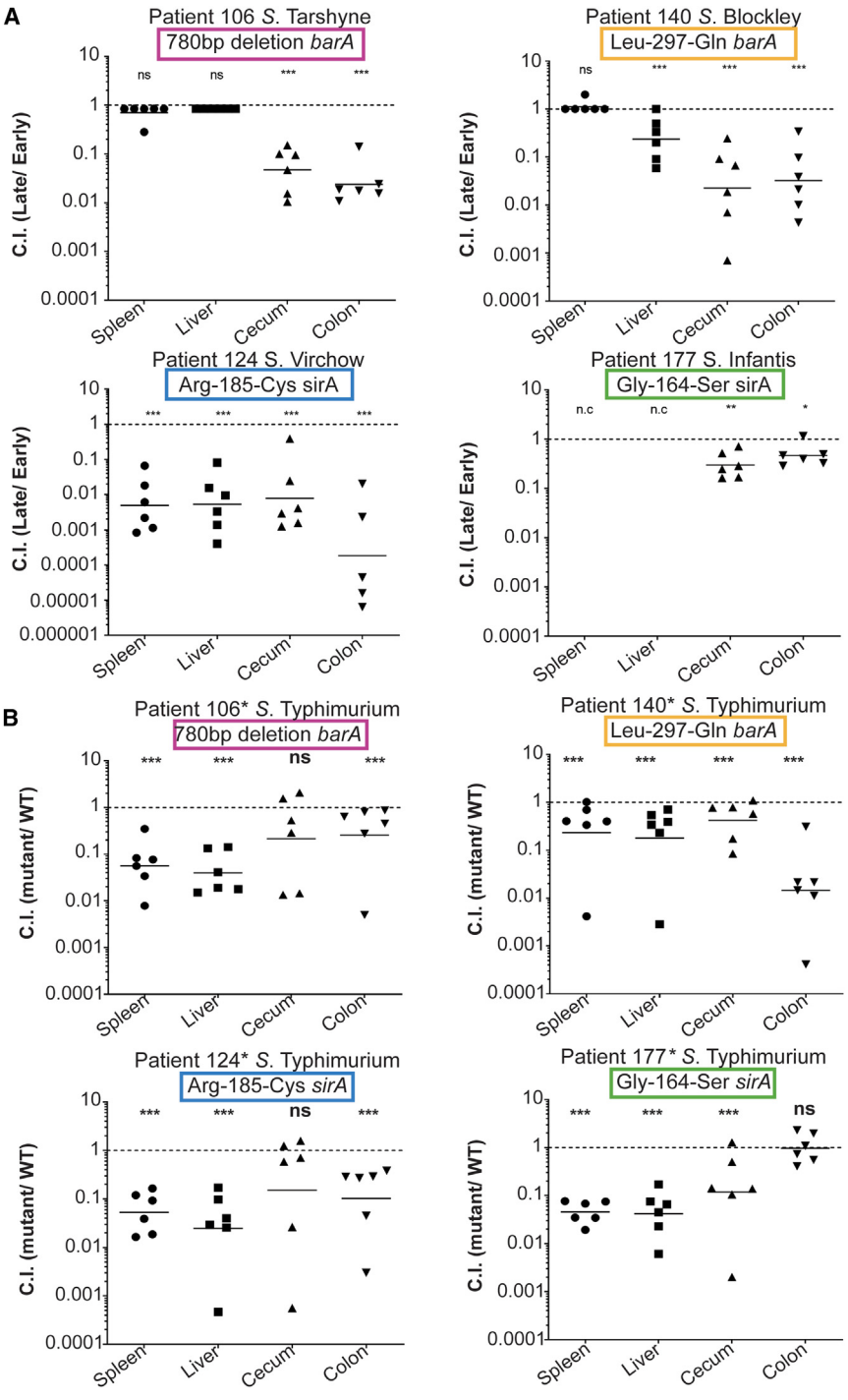


Figure 4. Mutations in *barA* and *sirA* acquired during persistence corresponds to attenuated *Salmonella* virulence

(A) Competitive index (CI) of late vs. early isolates with variants in *barA* or *sirA* after co-infection of C57BL/6 mice. CI was calculated for the spleen, liver, cecum, and colon of the mice and is plotted on the log scale. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ns, not significant; nc, no colonization was identified. The *Salmonella* serovar and location of the SNP in *barA* or *sirA* is shown at the top of each graph.

(B) Competitive infection in the acute mouse model using reconstructed *barA*/*sirA* mutations in *S.* Typhimurium SL1344. C57BL/6 mice were pretreated with streptomycin and infected with a mixture of *S.* Typhimurium SL1344 WT strain (Amp^r) and each one of the four *barA*/*sirA* mutants (Kan^r) identified in patients 106, 124, 140, and 177. At day 4 p.i., mice were euthanized and bacterial loads in the cecum, colon, liver, and spleen were determined by plating their homogenates on selective plates supplemented with ampicillin or kanamycin. Competitive index (CI) was calculated as $(mutant/WT)_{output}/(mutant/WT)_{input}$ and is plotted on the log scale. Each dot represents the CI in one mouse and the geometrical mean is shown by a horizontal bar. A competitive value of one indicates no significant difference in colonization between both strains.

the relative bacterial loads of late vs. early isolates. For each pair, the later isolate (carrying a *barA* or *sirA* mutation) was recovered from the intestinal sites at significantly lower quantities than the early isolate, suggesting that these *barA*/*sirA* mutations attenuate virulence in this mouse model by reducing the ability of the strain to colonize the gastrointestinal tract and in some cases systemic sites as well (Figure 4). Noteworthily, some of these serovars (e.g., *S.* Infantis) do not cause systemic infection in this mouse model; therefore, the role of *barA*/*sirA* mutations in systemic site colonization could not be assessed in these backgrounds.

Because the above analyses were done with the original clinical isolates that presented *barA*/*sirA* mutations in varying genetic backgrounds (serovars), we wanted to remove the influence of genetic background to understand the effect of the mutations more fully. We therefore reconstructed these mutations in an identical background of *S.* Typhimurium SL1344, using a scarless mutagenesis approach.[48] Quantitative reverse-transcription PCR (RT-qPCR) confirmed that the reconstructed *barA*/*sirA* mutants had decreased (10-to-100-fold) expression of SPI-1 (*hilA*, *invA*, *prgJ*, *invF*, *sipB*, and *sopB*) and SPI-4 (*siiA*)-associated genes (Figure S5). Similarly, the reconstructed mutants were recovered at significantly lower quantities at systemic sites (spleen and liver) than the wild-type (WT) strain in C57BL/6 mice. With one exception (patient 177 *sirA* mutation), the mutants also showed significantly lower bacterial load in the colon of infected mice, 4 days p.i. relative to the WT background (Figure 4B). Collectively, we concluded from these experiments that mutations in the BarA/SirA pathway lead to a decreased expression of SPI-1 and SPI-4 genes and virulence attenuation in an acute salmonellosis mouse infection.

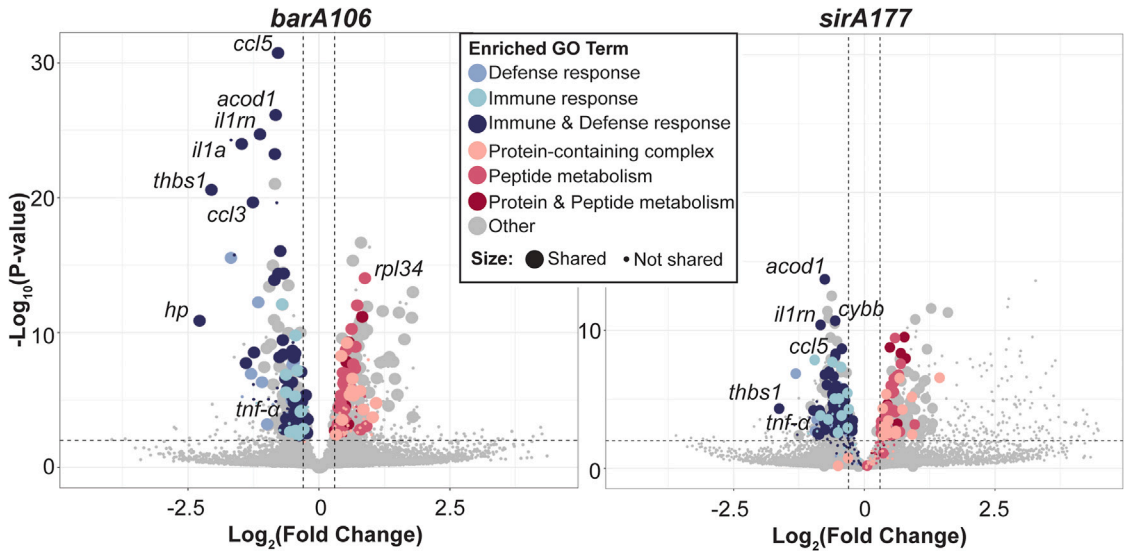


Figure 5. BMDMs infected with *barA/sirA* mutants provoke a weaker immune response

Volcano plots of DEGs between C57BL/6 BMDMs infected with either *S.* Typhimurium WT strain and isogenic *barA/sirA* mutants at MOI of 2.5. Larger point size indicates the gene is differentially expressed in both mutants; color indicates gene annotation with an enriched GO term.

Acquired mutations in barA and sirA provoke a weaker immune response in host cells

We hypothesized that genetic variation in *barA/sirA* acquired during *Salmonella* persistence in humans may affect the host response against *Salmonella* infection. To test this hypothesis, we infected bone-marrow-derived macrophages (BMDMs) from C57BL/6 mice with 3 × 10^6 colony-forming units (CFUs) of *S.* Typhimurium WT or an isogenic *barA/sirA* mutant strain at MOI of 2.5 (bacteria/macrophage). At 16 h p.i., total RNA was extracted, sequenced, and differentially expressed genes (DEGs) in BMDMs infected with WT and each *barA/sirA* mutant strain were determined by RNA-seq. The mutations from patient 106 (780 bp deletion in *barA*) and patient 177 (Gly-164-Ser in *sirA*) had the most marked differences in BMDM gene expression, particularly in genes involved in immune response (Figure 5), e.g., *il1A* (interleukin-1α), *il1rn* (interleukin 1 receptor antagonist), *acod1* (involved in mediating itaconate production, oxidative stress, and antigen processing), TNF-α (tumor necrosis factor alpha), *ccl3* (macrophage inflammatory protein 1α), and *ccl5* (involved in maintaining inflammation) (Table S7). Among down-regulated DEGs, GO terms including immune response, defense response, and response to cytokines were among the most significantly enriched (FDRs of 4.88E−38 and 2.48E−23, 8.03E−38, and 1.80E−21, and 1.16E−31 and 8.23E−23 for patients 106 and 177, respectively). Among the upregulated genes, GO terms such as protein-containing complex assembly, peptide metabolic process, and ribosome biogenesis were among the most significantly enriched (FDRs of 7.9E−11 and 0.030, 1.94E−26 and 2.09E−07, and 4.38E−08 and 0.0034). Taken together, these results indicate better growth of host cells infected with the mutant isolates (Table S8).

Although not reaching statistical significance, DEGs in BMDMs infected with strains from patients 140 and 124 showed similar expression patterns; of the 293 DEGs observed in both patients 106 and 177, 85% and 99% (249 and 291 genes) were differentially expressed in the same direction in patients 140 and 124, respectively (Table S7). We concluded from these results that the acquired mutations in the BarA/SirA pathway lead to reduced transcription of genes involved in the host immune response against *Salmonella* infection, most likely due to the suppressed expression of SPI-1 genes.

Acquired mutations in barA and sirA can cause long-term infections and prolonged shedding in a persistence mouse model

Our overarching hypothesis is that the milder immune response against *barA/sirA* mutants shown by the BMDM transcriptome experiment might allow these mutants to persist for a long time in their host. To test this hypothesis, we individually infected 7-week old CBA/CA mice, previously used as a *Salmonella* persistence model,[49–51] with 1 × 10^7 CFUs of *S.* Typhimurium WT or with one of its four isogenic *barA/sirA* mutants. At an early time point (day 6 p.i.), three of the four mutants showed lower levels of shedding compared with the WT strain, suggesting that these *barA/sirA* mutants had overall lower intestinal colonization. However, by later time points (days 9, 12, and 16 p.i.), all four mutants showed comparable levels of shedding to the WT strain, suggesting that, at later stages of infection, there were comparable levels of colonization across all strains, including WT (Figure 6A). In agreement with this finding, when mice were euthanized at day 21 p.i., similar bacterial loads of the four *barA/sirA* mutants and the WT were recovered from all body sites tested, except in the cecum, where one of the *barA* mutants (780 bp deletion, patient 106) showed a roughly 10-fold higher bacterial burden than the WT background (Figure 6B). Despite similar or higher levels of overall colonization by the *barA/sirA* mutants, RT-qPCR analysis of RNA extracted from the spleens of each of the *barA*- and WT-infected mice showed that the mutants elicited less IL-22 expression than WT (Figure 6C), in agreement with our BMDM results. Results from the persistence

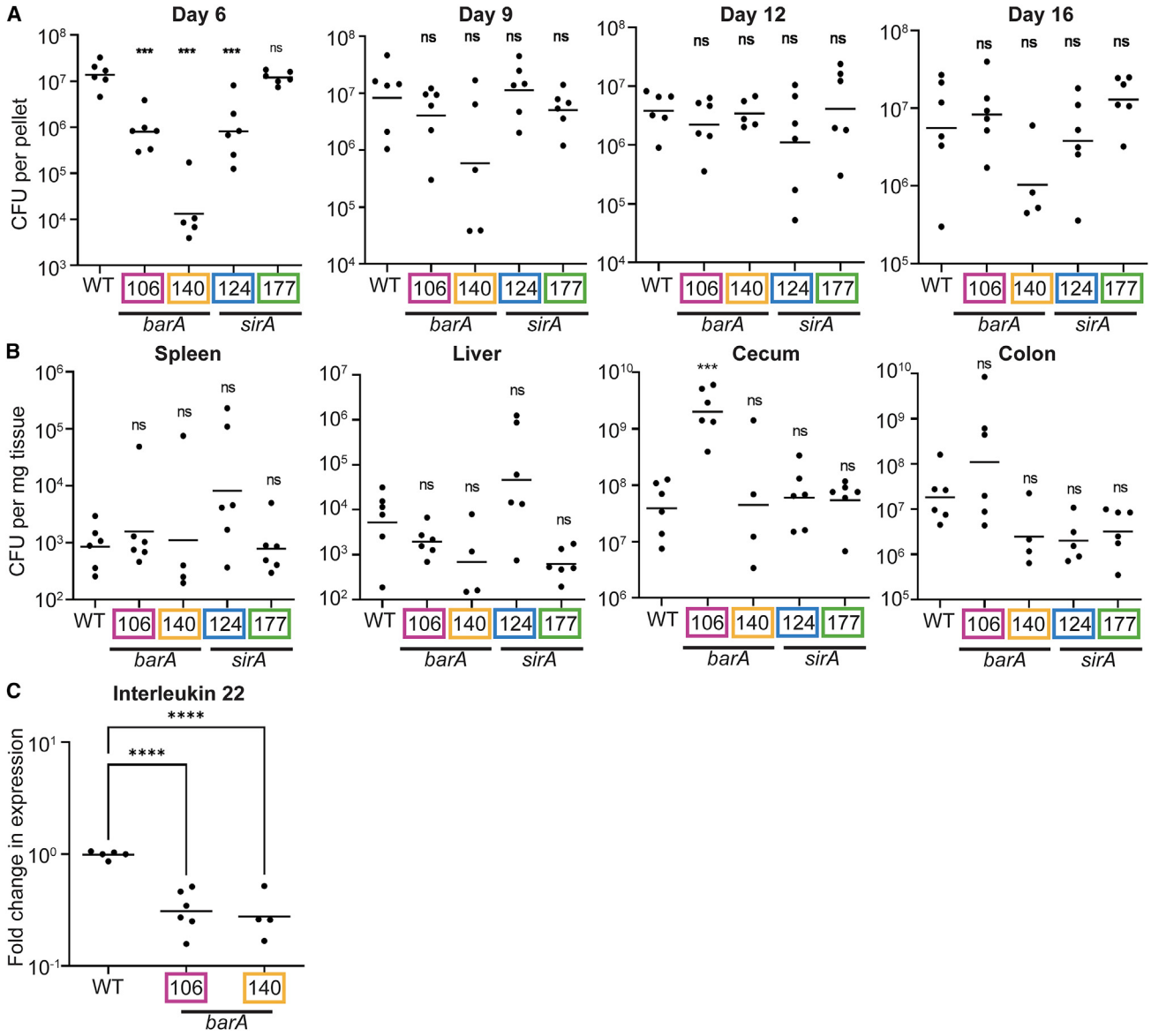


Figure 6. *barA*/*sirA* mutants can establish a long-term infection and prolonged shedding in a persistent mouse model

7-week-old CBA/CA mice were orally infected with *S.* Typhimurium WT or four isogenic *barA*/*sirA* mutant strains, named after the mutations found in isolates from patients (106, 140, 124, and 177).

(A) Mice feces were collected at days 6, 9, 12, and 16 p.i., and the number of CFUs per pellet is shown.

(B) *Salmonella* colonization at systemic (spleen and liver) and intestinal (cecum and colon) organs was determined at day 21 p.i. by plating tissue homogenates onto selective plates. Each dot represents the number of CFUs from one sample in one mouse and the geometrical mean is indicated by a horizontal bar.

(C) The expression level of interleukin 22 in the spleen of the infected mice was measured by RT-qPCR and normalized to the expression of the *hprt* housekeeping gene. Each dot shows the average of three RT-qPCR reactions from one mouse spleen sample, and the geometric mean of each sample is indicated by the horizontal line. One-way ANOVA was used to determine statistical significance. ****, $p < 0.0001$.

mouse model suggest a milder immune response in mice infected with *Salmonella* carrying mutations in the BarA/SirA pathway, despite colonization levels that were similar and sometimes exceeded those of the WT.

DISCUSSION

Using comparative genomics across a large set of persistent *Salmonella* isolates from diverse serovars, we identified convergent within-host mutations in global regulatory genes across nearly 50% of patients, including in the genes encoding the BarA/SirA TCRS. Using RNA-seq, we found that these mutations led to the downregulation of many genes encoded in SPI-1, SPI-4, and the genes encoding T3SS-1 associated effector proteins. Moreover, competition experiments in the acute salmonellosis mouse model revealed that the late isolates harboring a mutation in either *barA* or *sirA* were significantly less virulent than the early isolates. Using mouse BMDMs, we showed that the

transcriptional host response to these mutants was defined by a significant downregulation of genes involved in the immune, defense, and cytokine responses. In the persistence mouse model, we found that, at later time points, in contrast to the acute mouse model, *barA*/*sirA* mutants were colonizing intestinal and systemic sites and being shed at comparable or even higher levels to the WT strain.

Although our findings show convergent adaptation in BarA/SirA in persistent *Salmonella*, recent studies in other bacteria that cause persistent human infections have also identified within-host adaptations in global regulatory genes. In the context of the cystic fibrosis (CF) lung infections, mutations have been shown to accumulate in the genes encoding the (1) BarA/SirA ortholog, GacA/S, in persistent *Pseudomonas aeruginosa* infection,[52,53] (2) virulence transcriptional regulators AgrA and RsbU in persistent *Staphylococcus aureus* infection,[54] and (3) global regulatory proteins PhoQ, BigR, SpoT, and CpxA in *Achromobacter xylosoxidans* infection.[55] Outside of CF, (1) in a chronic diabetic foot infection, persistent *S. aureus* evolved mutations in the two-component regulatory systems, SsrA-SsrB and VraS/VraR, controlling aspects of *Staphylococcus aureus* pathogenicity,[56] and (2) in persistent *Mycobacterium abscessus* lung infection, this environmental pathogen evolved mutations in global regulators PhoR, Crp/Fnr, EngA, a TetR family member, and IdeR.[57] Taken together, our findings and these studies strongly suggest that positive selection of mutations in global regulatory pathways that control bacterial virulence may be a common path for diverse pathogens to transition from acute to persistent infection.

To understand the mechanism(s) by which mutations in key virulence regulators could lead to persistence in the human host, we profiled the transcriptional effects of mutations in the BarA/SirA two-component system carried by late isolates. We chose to focus on mutations that appeared in the BarA/SirA pathway due to their exceptionally high representation in many different patients and across multiple *Salmonella* serovars. We found that the genes downregulated in the persistent isolates of all four patients were strongly enriched for genes in SPI-1 and SPI-4. SPI-1 is known to control invasion of cells in the intestinal epithelium, leading to enteritis through a T3SS that is able to inject effector proteins directly into the cytoplasm of the host cells.[14] Moreover, SPI-1 effectors have previously been shown to be required for persistence in the mouse model.[58,59] SPI-4, which encodes the invasin protein SiiE, has been shown to be coregulated with SPI-1 and is believed to have a complementary role to SPI-1 during the enteric phase of *Salmonella* pathogenicity.[60] In contrast, we found far fewer shared upregulated genes in the persistent isolates, the only commonly upregulated gene being *ompF*, one of the major porins in *Salmonella*. It has recently been shown that *ompF* expression is increased when the T3SS-1 effector SseK1 is downregulated through OmpR, and that *Salmonella* Δ*sseK1* mutant strains have higher bile salt resistance and increased capacity to form biofilms. Therefore, it is possible that induced expression of *ompF* resulting from Ssek1 downregulation contributes to *Salmonella* persistence.[61]

To determine the host response to isolates with mutations in *barA*/*sirA*, we profiled the transcriptomes of infected BMDMs infected with WT or mutant isolates and found that these mutations led to a significant decrease in the host immune response, in particular the defense response and cytokine signaling. This may be the result of a downregulation of the SPI-1 T3SS needle inner rod component, PrgJ, that triggers the assembly of the NAIP/NLRC4 inflammasome, mediating caspase-1 activation, IL-1 family cytokine secretion, and pyroptosis of infected cells.[62] Indeed, when we infected *Salmonella*-resistant mice that permit long-term infection with *S.* Typhimurium, we found that, in contrast to the acute mouse model, *barA*/*sirA* mutants were shed and colonized intestinal and systemic sites of the persistence mouse model at comparable or even higher levels than the WT strain, despite displaying lower virulence in the acute mouse model.

Based on our results, we propose an evolutionary model by which mutations in global virulence regulators, such as the BarA/SirA TCRS (as well as other regulators), lead to the downregulation of virulence gene expression, including genes encoded on SPIs-1 and 4. Lower expression of these pathways at the early stage of the infection attenuates *Salmonella* pathogenicity and provokes a weaker immune response against such variants. These variants are then positively selected in the host and can persist longer while continuing to be shed to the environment.

Although unanswered questions remain about how bacterial pathogens establish and maintain persistent infections, this study provides key insights into the genetic adaptations underlying *Salmonella* persistence and the transcriptional mechanism by which they influence virulence. Not only does our large set of whole-genome sequences of persistent NTS of *Salmonella* provide an invaluable resource for the community, our study also serves as a template for how mechanisms of persistence can be studied in other bacteria. The size and scope of our study allowed us to pinpoint convergent evolution not only at the nucleotide and gene level, but also at the pathway level, where mutations are acquired in different genes that all share a common function. We believe that these findings of mutations in conserved regulatory pathways, leading to persistence, may also be relevant to other pathogens and have the potential to assist in the development of therapeutic approaches to prevent or treat persistent bacterial infections.

Limitations of the study

Although early isolates in our dataset were taken upon first culture-confirmed *Salmonella* diagnosis, we cannot be certain that all early samples were isolated at the initial onset of disease. Indeed, the presence in the blood of some early isolates may be indicative of the fact that these were taken at an advanced stage of infection, after *Salmonella* had already disseminated from the gut. Therefore, we may have missed some earlier transitions to persistence within patients. This may be why, in some patients, we do not find any variants between the early and late isolates. To this point, when we include patients with nonsynonymous or nonsense mutations in *barA* or *sirA* shared by the early and late isolate compared with the reference, we find that 416 isolates in 181 different patients present sequence variation in *barA*/*sirA*, including 24 isolates from 18 different patients with presumably inactivating mutations (insertions, deletions, or nonsense mutations) in *barA* or *sirA*. Because natural genetic variation in the *barA*/*sirA* genes is not high (Figure S2B), we think

that these cases may point to a transition to persistence having already occurred (Table S9). Thus, adaptations in *barA/sirA* may be even more prevalent in our dataset than we have conservatively calculated based on differences between early and late isolates. Additionally, we chose to use a reference genome against which to compare all isolate genomes, because we were interested in finding trends that occurred across the phylogeny and that were serovar-independent. Although this method proved to be fruitful, we could be missing additional strain or serovar-specific adaptations to a persistent lifestyle that are in genomic regions not present in the reference *S.* Typhimurium core genome. Lastly, an alternative mechanism of persistence could be that general dysregulation of virulence gene expression results in a global phenotypic heterogeneity that enables some bacteria to escape detection and killing by the host, which facilitates persistent infection. Although we chose to focus on *barA* and *sirA* mutations because they were the most prevalent in our dataset, future work could include comparative transcriptomics across other identified mutated regulators, which may reveal convergent transcriptional pathways and similar phenotypic outcomes, through signaling via the Rcs system[63] or DksA,[64] for example.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Human (clinical) bacterial isolates
  - Microbe strains
  - Animal models
  - Primary cell cultures
- METHOD DETAILS
  - Sample isolation
  - Whole genome sequencing and variant analysis
  - Construction of a core genome and estimation of *in vivo* mutation rate
  - Genome assembly, plasmid analysis, and antibiotic resistance analysis
  - Reconstruction of *barA*/*sirA* mutations in a *S.* Typhimurium background
  - Acute salmonellosis mouse model
  - Persistence mouse model experiments
  - Infection of bone-marrow derived macrophages
  - RNA extraction from early and persistent isolates with *barA*/*sirA* mutations
  - RNA extraction from spleen
  - Generation of RNA-Seq Data
  - RT-qPCR experiments
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Nucleotide diversity analysis
  - Analysis of RNA-Seq data

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.chom.2023.12.001.

ACKNOWLEDGMENTS

We acknowledge the Broad's Microbial 'Omics Core for generating the RNA-seq libraries and conducting preliminary analysis for all RNA-seq data, the Broad's Genomics Platform for sequencing, Rachel Newmiller, and the Broad's Bacterial Genomics group for helpful discussions. We are grateful to the National *Salmonella* Reference Center and Dr. Maya Davidovich-Cohen for routinely serotyping all clinical *Salmonella* isolates collected in Israel and to Dr. Alex Marzel for his previous epidemiological analysis, which set the foundation for this study. The work at the Gal-Mor laboratory was supported by the following grants: I-41-416.6-2018 from the German-Israeli Foundation for Scientific Research and Development (GIF), A128055 from the Research Cooperation Lower Saxony-Israel (The Volkswagen Foundation), and 2616/18 from the joint Israel Science Foundation (ISF)-Broad Institute program. The work at the Broad was supported by grants: 2616/18 from the joint ISF-Broad Institute program, U19AI110818 from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, and the Jane Coffin Childs Memorial Fund for Medical Research postdoctoral fellowship to A.G. Some figures were made in BioRender. The funders had no role in study design, data collection, and interpretation, or the decision to submit the work for publication.

AUTHOR CONTRIBUTIONS

O.G.-M., A.M.E., J.L., B.P., and A.G. conceived and designed the analysis. O.G.-M., A.M.E., J.L., B.P., B.A., and H.C. collected data for the study. O.G.-M., A.M.E., J.L., B.P., B.A., H.C., A.G., and A.L.M. contributed data or analysis tools. O.G.-M., J.L., B.P., B.A., H.C., and A.G. performed analysis. O.G.-M., B.P., and A.G. did the visualization. O.G.-M., A.M.E., J.L., B.P., A.G., and A.L.M. wrote and edited the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 26, 2022
Revised: October 4, 2023
Accepted: December 6, 2023
Published: December 29, 2023

REFERENCES

1. Cohen, N.R., Lobritz, M.A., and Collins, J.J. (2013). Microbial persistence and the road to drug resistance. Cell Host Microbe *13*, 632–642.
2. Issenhuth-Jeanjean, S., Roggentin, P., Mikoleit, M., Guibourdenche, M., de Pinna, E., Nair, S., Fields, P.I., and Weill, F.X. (2014). Supplement 2008–2010 (no. 48) to the White-Kauffmann-Le Minor scheme. Res. Microbiol. *165*, 526–530.
3. Gal-Mor, O. (2019). Persistent infection and long-term carriage of typhoidal and nontyphoidal salmonellae. Clin. Microbiol. Rev. *32*, e00088-18.
4. Crump, J.A., Luby, S.P., and Mintz, E.D. (2004). The global burden of typhoid fever. Bull. World Health Organ. *82*, 346–353.
5. Havelaar, A.H., Kirk, M.D., Torgerson, P.R., Gibb, H.J., Hald, T., Lake, R.J., Praet, N., Bellinger, D.C., de Silva, N.R., Gargouri, N., et al. (2015). World Health Organization global estimates and regional comparisons of the burden of foodborne disease in 2010. PLOS Med. *12*, e1001923.
6. Galán, J.E. (2001). *Salmonella* interactions with host cells: type III secretion at work. Annu. Rev. Cell Dev. Biol. *17*, 53–86.
7. Mills, D.M., Bajaj, V., and Lee, C.A. (1995). A 40 kb chromosomal fragment encoding *Salmonella typhimurium* invasion genes is absent from the

corresponding region of the *Escherichia coli* K-12 chromosome. Mol. Microbiol. *15*, 749–759.

8. Pavlova, B., Volf, J., Ondrackova, P., Matiasovic, J., Stepanova, H., Crhanova, M., Karasova, D., Faldyna, M., and Rychlik, I. (2011). SPI-1-encoded type III secretion system of *Salmonella enterica* is required for the suppression of porcine alveolar macrophage cytokine expression. Vet. Res. *42*, 16.
9. Zhao, X., Tang, X., Guo, N., An, Y., Chen, X., Shi, C., Wang, C., Li, Y., Li, S., Xu, H., et al. (2018). Biochanin A enhances the defense against *Salmonella enterica* infection through AMPK/ULK1/mTOR-mediated autophagy and extracellular traps and reversing SPI-1-dependent macrophage (MΦ) M2 polarization. Front. Cell. Infect. Microbiol. *8*, 318.
10. Sekirov, I., Gill, N., Jogova, M., Tam, N., Robertson, M., de Llanos, R., Li, Y., and Finlay, B.B. (2010). Salmonella SPI-1-mediated neutrophil recruitment during enteric colitis is associated with reduction and alteration in intestinal microbiota. Gut Microbes *1*, 30–41.
11. Winter, S.E., Thiennimitr, P., Winter, M.G., Butler, B.P., Huseby, D.L., Crawford, R.W., Russell, J.M., Bevins, C.L., Adams, L.G., Tsolis, R.M., et al. (2010). Gut inflammation provides a respiratory electron acceptor for Salmonella. Nature *467*, 426–429.
12. Ochman, H., Soncini, F.C., Solomon, F., and Groisman, E.A. (1996). Identification of a pathogenicity island required for Salmonella survival in host cells. Proc. Natl. Acad. Sci. USA *93*, 7800–7804.
13. Shea, J.E., Hensel, M., Gleeson, C., and Holden, D.W. (1996). Identification of a virulence locus encoding a second type III secretion system in Salmonella typhimurium. Proc. Natl. Acad. Sci. USA *93*, 2593–2597.
14. Lou, L., Zhang, P., Piao, R., and Wang, Y. (2019). Salmonella Pathogenicity Island 1 (SPI-1) and its complex regulatory network. Front. Cell. Infect. Microbiol. *9*, 270.
15. Martínez, L.C., Yakhnin, H., Camacho, M.I., Georgellis, D., Babitzke, P., Puente, J.L., and Bustamante, V.H. (2011). Integration of a complex regulatory cascade involving the SirA/BarA and Csr global regulatory systems that controls expression of the Salmonella SPI-1 and SPI-2 virulence regulons through HilD. Mol. Microbiol. *80*, 1637–1656.
16. Behlau, I., and Miller, S.I. (1993). A PhoP-repressed gene promotes Salmonella typhimurium invasion of epithelial cells. J. Bacteriol. *175*, 4475–4484.
17. Ellermeier, J.R., and Slauch, J.M. (2007). Adaptation to the host environment: regulation of the SPI1 type III secretion system in Salmonella enterica serovar Typhimurium. Curr. Opin. Microbiol. *10*, 24–29.
18. Teplitski, M., Goodier, R.I., and Ahmer, B.M.M. (2003). Pathways leading from BarA/SirA to motility and virulence gene expression in *Salmonella*. J. Bacteriol. *185*, 7257–7265.
19. Heeb, S., and Haas, D. (2001). Regulatory roles of the GacS/GacA two-component system in plant-associated and other Gram-negative bacteria. Mol. Plant. Microbe Interact. *14*, 1351–1363.
20. Foster, N., Tang, Y., Berchieri, A., Geng, S., Jiao, X., and Barrow, P. (2021). Revisiting persistent *Salmonella* infection and the carrier state: what do we know? Pathogens *10*, 1299.
21. J Barton, A., Hill, J., J Blohmke, C., and J Pollard, A. (2021). Host restriction, pathogenesis and chronic carriage of typhoidal Salmonella. FEMS Microbiol. Rev. *45*, fuab014.
22. Marzel, A., Desai, P.T., Goren, A., Schorr, Y.I., Nissan, I., Porwollik, S., Valinsky, L., McClelland, M., Rahav, G., and Gal-Mor, O. (2016). Persistent infections by nontyphoidal *Salmonella* in humans: epidemiology and genetics. Clin. Infect. Dis. *62*, 879–886.
23. Buchwald, D.S., and Blaser, M.J. (1984). A review of human salmonellosis: II. Duration of excretion following infection with Nontyphi Salmonella. Rev. Infect. Dis. *6*, 345–356.
24. Dixon, J.M. (1965). Effect of antibiotic treatment on duration of excretion of Salmonella typhimurium by children. Br. Med. J. *2*, 1343–1345.
25. Musher, D.M., and Rubenstein, A.D. (1973). Permanent carriers of nontyphosa salmonellae. Arch. Intern. Med. *132*, 869–872.
26. Vogelsang, Th.M., and Bøe, J. (1948). Temporary and chronic carriers of *Salmonella typhi* and *Salmonella paratyphi* B.J. Hyg. J. Hyg. (Lond) *46*, 252–261.
27. Stecher, B., and Hardt, W.D. (2011). Mechanisms controlling pathogen colonization of the gut. Curr. Opin. Microbiol. *14*, 82–91.
28. Octavia, S., Wang, Q., Tanaka, M.M., Sintchenko, V., and Lan, R. (2015). Genomic variability of serial human isolates of Salmonella enterica serovar Typhimurium associated with prolonged carriage. J. Clin. Microbiol. *53*, 3507–3514.
29. Kim, D., Song, L., Breitwieser, F.P., and Salzberg, S.L. (2016). Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. *26*, 1721–1729.
30. Merselis, J.G., Kaye, D., Connolly, C.S., and Hook, E.W. (1964). Quantitative bacteriology of the typhoid carrier state. Am. J. Trop. Med. Hyg. *13*, 425–429.
31. Murase, T., Yamada, M., Muto, T., Matsushima, A., and Yamai, S. (2000). Fecal excretion of Salmonella enterica Serovar typhimurium following a food-borne outbreak. J. Clin. Microbiol. *38*, 3495–3497.
32. Robertson, J., and Nash, J.H.E. (2018). MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. Microb. Genom. *4*.
33. Alcock, B.P., Raphenya, A.R., Lau, T.T.Y., Tsang, K.K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.-L.V., Cheng, A.A., Liu, S., et al. (2020). CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. *48*, D517–D525.
34. Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S., and Phillippy, A.M. (2016). Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. *17*, 132.
35. Platt, D.J., Taggart, J., and Heraghty, K.A. (1988). Molecular divergence of the serotype-specific plasmid (pSLT) among strains of *Salmonella typhimurium* of human and veterinary origin and comparison of pSLT with the serotype specific plasmids of S. enteritidis and S. Dublin. J. Med. Microbiol. *27*, 277–284.
36. Cohen, E., Rahav, G., and Gal-Mor, O. (2020). Genome sequence of an emerging Salmonella enterica serovar Infantis and genomic comparison with other S. Infantis strains. Genome Biol. Evol. *12*, 151–159.
37. Nadin-Davis, S., Pope, L., Chmara, J., Duceppe, M.O., Burke, T., Devenish, J., Andrievskaia, O., Allain, R., and Ogunremi, D. (2020). An unusual Salmonella enteritidis Strain carrying a modified virulence plasmid lacking the prot6e gene represents a geographically widely distributed lineage. Front. Microbiol. *11*, 1322.
38. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.
39. Urrutia, I.M., Fuentes, J.A., Valenzuela, L.M., Ortega, A.P., Hidalgo, A.A., and Mora, G.C. (2014). Salmonella Typhi shdA: pseudogene or allelic variant? Infect. Genet. Evol. *26*, 146–152.
40. Hamed, S., Wang, X., Shawky, R.M., Emara, M., Aldridge, P.D., and Rao, C.V. (2019). Synergistic action of SPI-1 gene expression in Salmonella enterica serovar typhimurium through transcriptional crosstalk with the flagellar system. BMC Microbiol. *19*, 211.
41. Ibarra, J.A., Knodler, L.A., Sturdevant, D.E., Virtaneva, K., Carmody, A.B., Fischer, E.R., Porcella, S.F., and Steele-Mortimer, O. (2010). Induction of Salmonella pathogenicity island 1 under different growth conditions can affect Salmonella-host cell interactions in vitro. Microbiology (Reading) *156*, 1120–1133.
42. Kim, K., Golubeva, Y.A., Vanderpool, C.K., and Slauch, J.M. (2019). Oxygen-dependent regulation of SPI1 type three secretion system by small RNAs in Salmonella enterica serovar Typhimurium. Mol. Microbiol. *111*, 570–587.
43. Martínez-Flores, I., Pérez-Morales, D., Sánchez-Pérez, M., Paredes, C.C., Collado-Vides, J., Salgado, H., and Bustamante, V.H. (2016). In silico

clustering of Salmonella global gene expression data reveals novel genes co-regulated with the SPI-1 virulence genes through HilD. Sci. Rep. *6*, 37858.

44. Singer, H.M., Erhardt, M., and Hughes, K.T. (2013). RflM functions as a transcriptional repressor in the autogenous control of the *Salmonella flagellar* master operon flhDC. J. Bacteriol. *195*, 4274–4282.
45. Singer, H.M., Kühne, C., Deditius, J.A., Hughes, K.T., and Erhardt, M. (2014). The Salmonella Spi1 virulence regulatory protein HilD directly activates transcription of the flagellar master operon flhDC. J. Bacteriol. *196*, 1448–1457.
46. Shelton, C.D., Yoo, W., Shealy, N.G., Torres, T.P., Zieba, J.K., Calcutt, M.W., Foegeding, N.J., Kim, D., Kim, J., Ryu, S., et al. (2022). Salmonella enterica serovar Typhimurium uses anaerobic respiration to overcome propionate-mediated colonization resistance. Cell Rep. *38*, 110180.
47. Palmer, A.D., and Slauch, J.M. (2017). Mechanisms of Salmonella pathogenesis in animal models. Hum. Ecol. Risk Assess. *23*, 1877–1892.
48. Hoffmann, S., Schmidt, C., Walter, S., Bender, J.K., and Gerlach, R.G. (2017). Scarless deletion of up to seven methyl-accepting chemotaxis genes with an optimized method highlights key function of CheM in Salmonella Typhimurium. PLoS ONE *12*, e0172630.
49. Katribe, E., Bogomolnaya, L.M., Wingert, H., and Andrews-Polymenis, H. (2009). Subspecies IIIa and IIIb Salmonellae are defective for colonization of murine models of salmonellosis compared to Salmonella enterica subsp. I serovar typhimurium. J. Bacteriol. *191*, 2843–2850.
50. Plant, J.E., Higgs, G.A., and Easmon, C.S. (1983). Effects of antiinflammatory agents on chronic Salmonella typhimurium infection in a mouse model. Infect. Immun. *42*, 71–75.
51. Weening, E.H., Barker, J.D., Laarakker, M.C., Humphries, A.D., Tsolis, R.M., and Bäumler, A.J. (2005). The Salmonella enterica serotype Typhimurium lpf, bcf, stb, stc, std, and sth fimbrial operons are required for intestinal persistence in mice. Infect. Immun. *73*, 3358–3366.
52. Bartell, J.A., Sommer, L.M., Haagensen, J.A.J., Loch, A., Espinosa, R., Molin, S., and Johansen, H.K. (2019). Evolutionary highways to persistent bacterial infection. Nat. Commun. *10*, 629.
53. Marvig, R.L., Sommer, L.M., Molin, S., and Johansen, H.K. (2015). Convergent evolution and adaptation of Pseudomonas aeruginosa within patients with cystic fibrosis. Nat. Genet. *47*, 57–64.
54. Long, D.R., Wolter, D.J., Lee, M., Precit, M., McLean, K., Holmes, E., Penewit, K., Waalkes, A., Hoffman, L.R., and Salipante, S.J. (2021). Polyclonality, shared strains, and convergent evolution in chronic cystic fibrosis Staphylococcus aureus airway infection. Am. J. Respir. Crit. Care Med. *203*, 1127–1137.
55. Khademi, S.M.H., Gabrielaite, M., Paulsson, M., Knulst, M., Touriki, E., Marvig, R.L., and Påhlman, L.I. (2021). Genomic and phenotypic evolution of Achromobacter xylosoxidans during chronic airway infections of patients with cystic fibrosis. mSystems *6*, e0052321.
56. Lavigne, J.P., Hosny, M., Dunyach-Remy, C., Boutet-Dubois, A., Schuldiner, S., Cellier, N., Yahiaoui-Martinez, A., Molle, V., La Scola, B., Marchandin, H., et al. (2021). Long-term intrahost evolution of Staphylococcus aureus among diabetic patients with foot infections. Front. Microbiol. *12*, 741406.
57. Bryant, J.M., Brown, K.P., Burbaud, S., Everall, I., Belardinelli, J.M., Rodriguez-Rincon, D., Grogono, D.M., Peterson, C.M., Verma, D., Evans, I.E., et al. (2021). Stepwise pathogenic evolution of *Mycobacterium abscessus*. Science *372*, eabb8699.
58. Kidwai, A.S., Mushamiri, I., Niemann, G.S., Brown, R.N., Adkins, J.N., and Heffron, F. (2013). Diverse secreted effectors are required for Salmonella persistence in a mouse infection model. PLoS ONE *8*, e70753.
59. Lawley, T.D., Bouley, D.M., Hoy, Y.E., Gerke, C., Relman, D.A., and Monack, D.M. (2008). Host transmission of Salmonella enterica serovar Typhimurium is controlled by virulence factors and indigenous intestinal microbiota. Infect. Immun. *76*, 403–416.
60. Kiss, T., Morgan, E., and Nagy, G. (2007). Contribution of SPI-4 genes to the virulence of *Salmonella enterica*. FEMS Microbiol. Lett. *275*, 153–159.
61. Hasan, M.K., Scott, N.E., Hays, M.P., Hardwidge, P.R., and El Qaidi, S. (2023). Salmonella T3SS effector SseK1 arginine-glycosylates the two-component response regulator OmpR to alter bile salt resistance. Sci. Rep. *13*, 9018.
62. Miao, E.A., Mao, D.P., Yudkovsky, N., Bonneau, R., Lorang, C.G., Warren, S.E., Leaf, I.A., and Aderem, A. (2010). Innate immune detection of the type III secretion apparatus through the NLRC4 inflammasome. Proc. Natl. Acad. Sci. USA *107*, 3076–3080.
63. Salvail, H., and Groisman, E.A. (2020). The phosphorelay BarA/SirA activates the non-cognate regulator RcsB in Salmonella enterica. PLOS Genet. *16*, e1008722.
64. Zhu, Y.Y., Machleder, E.M., Chenchik, A., Li, R., and Siebert, P.D. (2001). Reverse transcriptase template switching: a SMART™ approach for full-length cDNA library construction. BioTechniques *30*, 892–897.
65. Wang, R.F., and Kushner, S.R. (1991). Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in Escherichia coli. Gene *100*, 195–199.
66. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv.
67. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLOS One *9*, e112963.
68. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately maximum-likelihood trees for large alignments. PLOS One *5*, e9490.
69. Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. *49*, W293–W296.
70. Didelot, X., and Wilson, D.J. (2015). ClonalFrameML: efficient inference of recombination in whole bacterial genomes. PLOS Comput. Biol. *11*, e1004041.
71. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. *19*, 455–477.
72. Robertson, J., Bessonov, K., Schonfeld, J., and Nash, J.H.E. (2020). Universal whole-sequence-based plasmid typing and its utility to prediction of host range and epidemiological surveillance. Microb. Genom. *6*.
73. van Dijk, L.R., Walker, B.J., Straub, T.J., Worby, C.J., Grote, A., Schreiber, H.L., Anyansi, C., Pickering, A.J., Hultgren, S.J., Manson, A.L., et al. (2022). StrainGE: a toolkit to track and characterize low-abundance strains in complex microbial communities. Genome Biol. *23*, 74.
74. Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R., Potter, S.C., Finn, R.D., and Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. *47*, W636–W641.
75. Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. *35*, 1547–1549.
76. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.
77. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
78. Azriel, S., Goren, A., Rahav, G., and Gal-Mor, O. (2016). The stringent response regulator DksA is required for Salmonella enterica serovar Typhimurium growth in minimal medium, motility, biofilm formation, and intestinal colonization. Infect. Immun. *84*, 375–384.
79. Boichis, E., Ran Sapir, S., and Herskovits, A.A. (2022). Bone marrow-derived macrophage (BMDM) infection by Listeria monocytogenes. Methods Mol. Biol. *2427*, 83–93.

80. Bhattacharyya, R.P., Bandyopadhyay, N., Ma, P., Son, S.S., Liu, J., He, L.L., Wu, L., Khafizov, R., Boykin, R., Cerqueira, G.C., et al. (2019). Simultaneous detection of genotype and phenotype enables rapid and accurate antibiotic susceptibility determination. Nat. Med. *25*, 1858–1864.

81. Shishkin, A.A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., Chen, J., Bhattacharyya, R.P., Rudy, R.F., Patel, M.M., et al. (2015). Simultaneous generation of many RNA-seq libraries in a single reaction. Nat. Methods *12*, 323–325.

82. Cohen, H., Adani, B., Cohen, E., Piscon, B., Azriel, S., Desai, P., Bähre, H., McClelland, M., Rahav, G., and Gal-Mor, O. (2022). The ancestral stringent response potentiator, DksA has been adapted throughout *Salmonella* evolution to orchestrate the expression of metabolic, motility, and virulence pathways. Gut Microbes *14*, 1997294.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial strains		
Patient 106 early isolate	This study	*Salmonella enterica* serovar Tarshyne isolate 101101
Patient 106 late isolate	This study	*Salmonella enterica* serovar Tarshyne isolate 101626 (780bp deletion in *barA*)
Patient 124 early isolate	This study	*Salmonella enterica* serovar Virchow isolate 100296
Patient 124 late isolate	This study	*Salmonella enterica* serovar Virchow isolate 100828 (Arg-185-Cys *sirA*)
Patient 140 early isolate	This study	*Salmonella enterica* serovar Blockley isolate 146086
Patient 140 late isolate	This study	*Salmonella enterica* serovar Blockley isolate 146681 (Leu-297-Gln *barA*)
Patient 177 early isolate	This study	*Salmonella enterica* serovar Infantis isolate 115869
Patient 177 late isolate	This study	*Salmonella enterica* serovar Infantis isoalte117169 (Gly-164-Ser *sirA*)
S. Typhimurium SL1344	*Salmonella* genetic stock center	wild type Sm^r *xyl hisG rpsL*
S. Typhimurium SL1344 780bp deletion (p106) *barA*	This study	*S.* Typhimurium SL1344 780bp deletion in *barA*
S. Typhimurium SL1344 Arg-185-Cys (p124) *sirA*	This study	*S.* Typhimurium SL1344 Arg-185-Cys substitution in *sirA*
S. Typhimurium SL1344 Leu-297-Gln (p140) *barA*	This study	*S.* Typhimurium SL1344 Leu-297-Gln substitution in *barA*
S. Typhimurium SL1344Gly-164-Ser (p177) *sirA*	This study	*S.* Typhimurium SL1344 Gly-164-Ser substitution in *sirA*
Deposited Data		
WGS and RNA-Seq data	This study	BioProject ID: PRJNA847966
Experimental models: Cell cultures		
BMDMs were isolated from the femur leg bone of 7-week-old C57BL/6 mice	C57BL/6 female mice	Isolated from the femur leg bone of 7-week-old mice.
Experimental models: Organisms/strains		
C57BL/6 mice	Envigo, Israel	Homozygous for a mutant *Nramp1* allele, $Nramp1^{G169D}$, develop lethal infections and were used for acute salmonellosis mouse model
CBA/CA mice	Envigo, Israel	Encode wildtype *Nramp1* and were used for prolonged *Salmonella* infection model
Oligonucleotides		
barA patient 106 scarless F GCTGATCTTTGGCTGGCGGCTTATGCGCGA TGTCACCGGG AGGGTTTTCCCAGTCACGAC	This Study	For *barA* patient 106 deletion
barA patient106 scarless R GGCGAGAAATACTGGCATCGGCCTGGCGA AACGCCTGAAA TGCTTCCGGCTCGTATGTTG	This study	
barA patient106 TC2 F TGGATCTCAAGTCGGTCAGG	This study	

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
barA patient106 TC2 R GGCGAGAAATACTGGCATCGGCCTGGCGAAAC GCCTGAAACCCGGTGACATCGCGCATAA	This study	
sirA patient124 scarless F GTCCTAAAACGGTGAACAGCTATTGCTATCGT ATGTTCAG AGGGTTTTCCCAGTCACGAC	This study	For *sirA* patient 124 mutation
sirA patient124 scarless R CAGGTGAGTCAGCTCAACATCACCATGAATGTT TAATTTA TGCTTCCGGCTCGTATGTTG	This study	
sirA patient124 TC2 F GTCCTAAAACGGTGAACAGCTATT GCTATCGTATGTTCAG	This study	
barA patient140 scarless F AGCGGCGCGTATTAAGTCGGAGTTCC**A**GGCG AACATGTCG AGGGTTTTCCCAGTCACGAC	This study	For *barA* patient 140 mutation
barA patient140 scarless R TAAAGCCAATGACGCCGTTCAGCGGCGTTC GCAGTTCGTG TGCTTCCGGCTCGTATGTTG	This study	
barA patient140 TC2 F AGCGGCGCGTATTAAGTCGGAGTT CC**A**GGCGAACATGTC	This study	
barA patient140 TC2 R GCGGAACGCTCAATGGTGTT	This study	
sirA patient177 scarless F TATGCTGATGATCACCAAGAGTCAGAAGGT CAATGAGATT AGGGTTTTCCCAGTCACGAC	This study	For *sirA* patient 177 mutation
sirA patient177 scarless R AGCTGTTCACCGTTTTAGGACTGAGATTCAG CTGTTCTGA TGCTTCCGGCTCGTATGTTG	This study	
sirA patient177 TC2 F TATGCTGATGATCACCAAGAGTCA GAAGGTCAATGAGATT	This Study	
IL- 22 RT F TGGCTCTGCAGGATTTTCATG	This study	RT-PCR of IL-22
IL- 22 RT R CGCCTTGATCTCTCCACTCT	This study	
hprt RT F AGTGTTGGATACAGGCCAGAC	This study	RT-PCR of *hprt*
hprt RT R CGTGATTCAAATCCCTGAAGT	This study	
16S rRNA RT F GGTTAAGTCCCGCAACGAG	This study	RT-PCR of 16S rRNA
16S rRNA RT R CTTCTCTTTGTATGCGCCATTG	This study	
hilA RT F TACAGCAAGCGCTTAAATTGCT	This study	RT-PCR of *hilA*
hilA RT R ATCCCCATTTGCGCCATGCT	This study	
invF RT F GGGAGAAGACTATGGCGTTTCT	This study	RT-PCR of *invF*
invF RT R GCAGCGATTGCGCCATACG	This study	
invA RT F TCCGCTAATTTGATGGATCTCA	This study	RT-PCR of *invA*
invA RT R ATCCGGAAAACGACCTTCAATC	This study	

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
prgJ RT F ATTGTCTCGCTGGATGACCG	This study	RT-PCR of *prgJ*
prgJ RT R CCGTCACCAGATTAGGGTCC	This study	
sopB RT F GGGAAARGGCGTATGCAGT	This study	RT-PCR of *sopB*
sopB RT R CCATGACGTATCCCGCAAA	This study	
sipB RT F ACTGGAGTCTCGTCTGGCG	This study	RT-PCR of *sipB*
sipB RT R GTCATAAACACTCTTGGCGGTA	This study	
siiA RT F CCTGATGTAGTTATTGACATGA	This study	RT-PCR of *siiA*
siiA RT R TCACTCTGRCACCTTTTTATTA	This study	
Recombinant DNA		
pWRG730 Plasmid	Hoffmann et al. [48]	pSIM5 derivative, temperature-sensitive replication (30°C) and Red recombinase expression (42°C), Tet-inducible expression of I-SceI, Cm^r
pWRG717 Plasmid	Hoffmann et al. [48]	pBluescript II SK+ derivative, aph resistance cassette and I-SceI cleavage site, Km^r, Amp^r
pWSK129 Plasmid	Wang and Kushner[65]	Km^r low copy number cloning vector
pWSK29 Plasmid	Wang and Kushner[65]	Amp^r low copy number cloning vector
Software and algorithms		
BWA mem (version 0.7.12)	Li[66]	https://github.com/lh3/bwa
Pilon (version 1.12)	Walker et al. [67]	https://github.com/broadinstitute/pilon
Fasttree (version 2.1.14)	Price et al. [68]	https://anaconda.org/bioconda/fasttree
itol (version 6.5.6)	Letunic and Bork[69]	https://itol.embl.de/
ClonalFrameML (version 1.12)	Didelot and Wilson[70]	https://github.com/xavierdidelot/ClonalFrameML
SPADES (version 3.1.1)	Bankevich et al. [71]	https://cab.spbu.ru/files/release3.1.1/manual.html
GAEMR (version 1.0.1)	Robertson et al. [72]	https://software.broadinstitute.org/software/gaemr/
StrainGE	van Dijk et al. [73]	https://github.com/broadinstitute/StrainGE
MUSCLE	Madeira[74]	https://www.ebi.ac.uk/Tools/msa/muscle/
MEGA X	Kumar et al.[75]	https://www.megasoftware.net/
edgeR (version 3.32.1)	Robinson et al.[76,77]	https://bioconductor.org/packages/release/bioc/html/edgeR.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and materials should be directed to and will be fulfilled by the lead contact, Dr. Ohad Gal-Mor, Ohad.Gal-Mor@sheba.health.gov.il.

Materials availability

All mutants constructed in this study are available to the community upon request.

Data and code availability

- All sequencing files from this study are available at the Sequence Read Archive [PRJNA847966].
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human (clinical) bacterial isolates

From one of the largest retrospective studies of *Salmonella* persistence, including 48,345 culture-confirmed salmonellosis cases that occurred in Israel over 17 years (between 1995 and 2012),[22] we assembled a collection of 639 longitudinal *S. enterica* isolates representing 49 NTS serovars obtained during persistent infections from 256 patients. Each patient was represented by 2-5 independent isolates obtained during different stages of infection that lasted between 30 and 2,001 days from the first culture-confirmed diagnosis. All patients' samples for whom two or more longitudinal isolates were kept as glycerol frozen stock, were selected for WGS. Sample collection was performed under approval number 5402-18-SMC of the Sheba Medical Center ethics committee.

Microbe strains

All clinical sequenced isolates and their metadata are listed in Table S1. *S.* Typhimurium SL1344 was used to reconstruct the *barA*/*SirA* mutations found in the clinical isolates.

Animal models

All experiments in the mouse models were conducted according to the ethical requirements of the Animal Care Committee of the Sheba Medical Center (Approval # 1182/18 and 0009-23) and in line with the national guidelines. Female C57BL/6 mice (Envigo, Israel) were infected at an age of 7-10 weeks. For the persistent infection mouse model, seven-week-old CBA/CA female mice (Envigo, Israel) were used. Mice were hosted at the SPF animal facility of the Sheba Medical Center in enriched cages (5-6 mice per cage) and were given food and water *ad libitum*.

Primary cell cultures

BMDMs were isolated from the femur leg bone of 7-week old C57BL/6 female mice.

METHOD DETAILS

Sample isolation

Clinical isolates were plated on XLD agar plate, and a single colony was grown in LB overnight at 37°C under aerobic conditions on a Roller Drum. The next day the DNA was extracted from 500 μl of culture using the GeneElute bacterial genomic DNA kit (Sigma-Aldrich). gDNA was quantified, assessed for its quality and whole genome sequenced.

Whole genome sequencing and variant analysis

Purified DNA (0.2 ng) was used as input into a miniaturized version of the Nextera-XT Library Preparation Kit (Illumina Inc.). All reactions were scaled to one-fourth their original volumes. Libraries were constructed according to the manufacturer's instructions with the following minor modifications to boost the insert size, thereby reducing read overlap during paired-end sequencing: i) tagmentation time was reduced from 5 minutes to 1 minute; individual sample libraries were pooled at equimolar concentration. The final library pool underwent an additional size selection using a 0.7X SPRI using AMPure XP beads (Beckman Coulter) and was sequenced on a NovaSeq SP at 300 cycles for 2x150 pair-end reads. Genome coverage was approximately 150-fold, on average. Resulting fastq files were checked for quality using FastQC and checked for contamination using Centrifuge.[29] In total, eight isolates were removed due to non-*Salmonella* genomic content.

BWA mem (version 0.7.12)[66] was used to align each of the 639 genome sequences to the *S.* Typhimurium SL1344 reference genome (accession number NC_016810.1), and Pilon (version 1.12) was used to call variants between early and late isolates.[67] VCF files were filtered for variants containing the filter status PASS, and variants labeled as imprecise were removed. VCF files for same-patient isolates were then compared to find all variants that differed between early and late isolates. Variants between same-patient isolates and different-patient isolates were quantified, and a cutoff of 36 bp was chosen to distinguish a persistent infection from a potential reinfection based on the longest time of infection in our dataset and previously published SNP rates.[22] Using a SNP-based alignment, Fasttree (version 2.1.14) was used to construct the phylogenetic tree of *Salmonella* isolates.[68] Tree visualization was performed using itol (version 6.5.6).[69]

Construction of a core genome and estimation of *in vivo* mutation rate

To determine an *in vivo* mutation rate, we first filtered sites along the *S.* Typhimurium SL1344 reference alignment that were i) low coverage (<5 reads at a position for any sample) or ii) potentially recombined based on analysis with ClonalFrameML (version 1.12).[70] From this core alignment, the mutation rate was estimated from the number of SNPs between the first and last isolate for each patient (ignoring intermediate isolates for this analysis), divided by the number of days between the two isolates. We estimated the rate and the standard deviation using the *fitdistr* function in the "MASS" package, using the Poisson distribution. Only isolate pairs that had less than 36 SNPs were included in this analysis. Lambda was found to be 0.02627, with a standard deviation of 0.02664, equating to 9.589 SNPs per year.

Genome assembly, plasmid analysis, and antibiotic resistance analysis

Genomes were assembled using SPADES (version 3.1.1) on paired-end fastq files, using default settings.[71] GAEMR (version 1.0.1) (http://software.broadinstitute.org/software/gaemr/) was used to check for quality of assembly and to generate standard assembly files. Draft genome assemblies were annotated with VESPER.

MOB-suite was used on the WGS assemblies to predict potential plasmids and then to cluster them by similarity, reconstruct the plasmid sequences, and make *in silico* predictions of the replicon family, relaxase type, and transferability.[32,72] The Resistance Gene Identifier (RGI) was used to identify potential antibiotic resistance genes in the genome assemblies using the Comprehensive Antibiotic Resistance Database (CARD).[33]

Reconstruction of *barA/sirA* mutations in a *S.* Typhimurium background

Markerless and scarless mutations in the *barA* and *sirA* genes identical to the mutations found in the late isolates of patients 106, 140, 124, and 177 were reconstructed using a λ Red recombinase system and the temperature-sensitive plasmids pWRG730 and pWRG717 [Key Resources Table] according to Hoffmann et al.[48] All primers used to create the mutations are listed in the Key Resources Table. Successful recombinants were selected by the I-SceI expression system from pWRG730 on LB plates containing 10 μg/mL chloramphenicol and 500 ng/mL anhydrotetracycline. pWRG730 curing was conducted by plating on LB plates and incubation at 42°C overnight. All induced mutations were verified by sanger sequencing.

Acute salmonellosis mouse model

All experiments in the mouse models were conducted according to the ethical requirements of the Animal Care Committee of the Sheba Medical Center (Approval # 1182/18 and 0009-23) and in line with the national guidelines. Female C57BL/6 mice (Envigo, Israel) were infected at an age of 7-10 weeks as previously described.[78] Briefly, Streptomycin (20 mg per mouse) was given by oral gavage in saline 24 h prior to infection. Mice were infected by oral gavage with $\sim 4\times10^6$ bacteria as a mixed (1:1) inoculum containing either an early isolate genotype strain resistant to ampicillin (harboring pWSK29) and late isolate genotype strain resistant to kanamycin (harboring pWSK129) [Key Resources Table].[65] For competition experiments using the *barA/sirA* mutants in the *S.* Typhimurium background, mice were coinfected with a mixture of 7×10^6 CFU of *S.* Typhimurium SL1344 WT (carrying ampicillin resistance) and the *barA/sirA* mutants (carrying kanamycin resistance). A mixed inoculum of *S.* Typhimurium SL1344 strains carrying pWSK29 or pWSK129 was used as a positive control. This control demonstrated a CI value of 1 which indicates similar fitness in mice [Key Resources Table].

Four days post infection, mice were humanely euthanized, and tissues were harvested aseptically for bacterial enumeration. Tissues were collected on ice and homogenized in 700 μl of saline using the BeadBlaster24 microtube homogenizer (Benchmark scientific). Serial dilutions of the homogenates were plated on XLD agar plates supplemented with ampicillin or kanamycin, incubated overnight, and counted to calculate bacterial tissue burdens. The competitive index was calculated as $[\text{late isolate/early isolate}]_{output}/[\text{late isolate/early isolate}]_{input}$.

Persistence mouse model experiments

Seven-week-old CBA/CA female mice (Envigo, Israel) were pretreated with streptomycin (20 mg per mouse) 24 h prior to infection. Groups of 6 mice were infected with 1×10^7 CFU of WT *S.* Typhimurium SL1344 or one of its four *barA/sirA* isogenic mutants. Fresh mice feces were individually collected from each mouse at day 6, 9, 12, and 16 p.i. and homogenized in 700 μl saline using BeadBlaster24 microtube homogenizer (Benchmark scientific). Serial dilutions were plated on XLD agar plates supplemented with streptomycin (50 μg/ml) and the bacterial loads per pellet was determined. At day 21 p.i, mice were euthanized, and tissues were harvested aseptically, homogenized, and plated onto selective XLD plates for bacterial enumeration.

Infection of bone-marrow derived macrophages

BMDMs were isolated from the femur leg bone of 7-week old C57BL/6 female mice according to Boichis, E. et al.[79] Briefly, femur bones were incubated for 10 min in 5 ml BMDM medium [50% DMEM high glucose, 20% FBS, 30% L-929 conditioned medium, 2 mM L- glutamine, 1 mM sodium pyruvate, 50 nM β-mercaptoethanol], containing 1% penicillin/streptomycin. The bones were cut at their ends and bone marrow was washed out from the bones with 10 ml BMDM medium supplemented with 1% penicillin/streptomycin using a syringe equipped with a 27G needle. Bones extract was centrifuged for 1 min at 50 g at room temperature and the supernatant containing BMDMs was transferred to a new 15 ml conical tube and centrifuged at 125 g for 10 min. The pellet was resuspended in 10 ml BMDM medium with 1% penicillin/streptomycin and seeded in an N-TC 145/20 mm petri dish containing 20 ml pre-warmed BMDM medium with 1% penicillin/streptomycin and incubated for 7 days at 37°C under 5% CO_2 atmosphere. At day 3 post incubation, 10 ml of fresh BMDM medium with 1% penicillin/streptomycin were added to each plate. To harvest the cells, plates were washed with ice-cold PBS (without Mg^{2+}) and incubated for 10 min in 10 ml ice-cold PBS at 4°C. Scraped cells were centrifuged at 125 g for 10 min at room temperature and resuspend in BMDM medium without antibiotics.

BMDMs were seeded in a 6-well dish (1.25×10^6 cells/ per well) and infected with overnight-grown *Salmonella* cultures grown in LB at MOI of 2.5. Infected cells were centrifuged at 1000 x g for 5 minutes and incubated at 37°C for 30 minutes in a 5% CO_2 forced-air incubator. The plate was then washed 3 times with PBS and 3 ml of fresh warm BMDM medium supplemented with 100 μg/ml gentamicin was added to each well followed by an incubation at 37°C. After 90 min, the medium was removed and a fresh warm BMDM medium supplemented with 10 μg/ml gentamicin was added to each well followed by an overnight incubation at 37°C in a 5% CO_2

atmosphere. In the next morning, the plate was washed 3 times with PBS, and 400 μl of RNA SHILD (Zymo Research) was added to each well. Quick- DNA/RNA Viral MagBead (Zymo Research) kit was used to extract RNA from the cells according to the manufacturer's instructions, including TURBO DNase (Invitrogen) treatment, to diminish any genomic DNA contamination. 1 μl of SUPERase In-RNase Inhibitor (Thermo Fisher Scientific) was added to each sample.

RNA extraction from early and persistent isolates with *barA/sirA* mutations

Four patients with different *barA/sirA* mutations in the late isolate were chosen for comparative RNA-Seq analysis. Isolates were chosen due to their high genomic similarity between early and late same-patient isolates. Patients 106 and 140 had only a single *barA* deletion or SNP, respectively between the early and late isolates, while patient 124 had the *sirA* SNP and one other synonymous SNP in the hypothetical protein SL1344_RS10945, and patient 177 had the *sirA* SNP and one intergenic SNP 131 bp upstream of SL1344_RS12695, transaldolase A. Isolates were grown in 2 ml LB overnight at 37°C on a roller drum. The next day the bacteria were subcultured (1:100 dilution) into 2 ml fresh LB and grew in 15×150 mm glass tubes on a roller drum for 3 hr at 37°C. RNA was extracted from 500 μl cultures using the Qiagen RNA protect Bacteria Reagent and RNeasy mini kit according to the manufacturer's instructions, including an on-column DNase digest using the RNase free DNase set (Qiagen). The quantity and quality of the extracted RNA were determined by Nanodrop 2000c (Thermo Fisher Scientific). To diminish any genomic DNA contamination, RNA was re-treated with an RNase-free DNase (QIAGEN). Cell pellets resuspended in 0.5 mL Trizol reagent (ThermoFisher Scientific) were transferred to 2 mL FastPrep tubes (MP Biomedicals) containing 0.1 mm Zirconia/Silica beads (BioSpec Products) and bead beaten for 90 seconds at 10 m/sec speed using the FastPrep-24 5G (MP Biomedicals). After addition of 200 ul chloroform, each sample tube was mixed thoroughly by inversion, incubated for 3 minutes at room temperature, and spun down 15 minutes at 4°C. The aqueous phase was mixed with an equal volume of 100% ethanol, transferred to a Direct-zol spin plate (Zymo Research), and RNA was extracted according to the Direct-zol protocol (Zymo Research).

RNA extraction from spleen

Mice spleen tissues (7-20 mg) were added to 1.5 ml of RNAprotect Tissue Reagent (Qiagen) and kept at 4°C until processing. RNAprotect was replaced with 600 μl of buffer RLT of the RNeasy Plus Mini kit (Qiagen) and homogenized using hand homogenizer. RNA purification was carried out according to the manufacturer's instructions, including two cycles of TURBO DNase (Invitrogen) treatment. RT-qPCR for IL-22 was conducted using the primers IL- 22 RT F and IL- 22 RT R and the housekeeping hypoxanthine guanine phosphoribosyl transferase (HPRT) gene as a normalization control.

Generation of RNA-Seq Data

Illumina cDNA libraries were generated using a modified version of the RNAtag-seq protocol.[80,81] Briefly, 0.5-1 μg of total RNA was fragmented, depleted of genomic DNA, dephosphorylated, and ligated to DNA adapters carrying 5'-AN_8-3' barcodes of known sequence with a 5' phosphate and a 3' blocking group. Barcoded RNAs were pooled and depleted of rRNA using the RiboZero rRNA depletion kit (Illumina). Pools of barcoded RNAs were converted to Illumina cDNA libraries in 2 main steps: (i) reverse transcription of the RNA using a primer designed to the constant region of the barcoded adaptor with addition of an adapter to the 3' end of the cDNA by template switching using SMARTScribe (Clontech) as described[64]; (ii) PCR amplification using primers whose 5' ends target the constant regions of the 3' or 5' adaptors and whose 3' ends contain the full Illumina P5 or P7 sequences. cDNA libraries were sequenced to generate paired end reads.

RT-qPCR experiments

All primers used for RT-qPCR-analysis are listed in the Key Resources Table. RNA was extracted from subcultures of *S.* Typhimurium SL1344 and its isogenic *barA/sirA* mutants grown in LB to the late logarithmic phase in LB medium at 37°C, using the Qiagen RNAprotect Bacteria Reagent and RNeasy mini kit (Qiagen) according to the manufacturer's instructions, including an on-column DNase digest using the RNase free DNase set (Qiagen). To diminish any genomic DNA contamination, RNA was re-treated with TURBO DNase (Invitrogen). cDNA was synthesized from 200 ng of purified RNA, using qScript cDNA Synthesis Kit (Quanta-bio) and a T100 thermal cycler (Bio-Rad Laboratories). Real-Time PCR reactions were performed in a StepOne Real-Time PCR platform (Applied Biosystems). Each reaction was carried out using 1 μl of cDNA, in a total volume of 20 μl on a 96-well optical reaction plate (Applied Biosystems). 16S rRNA gene was used as endogenous normalization control. Calculation of $\Delta\Delta C_t$ and fold change in gene expression was calculated as previously explained.[82]

QUANTIFICATION AND STATISTICAL ANALYSIS

For all *in vivo* studies, statistical analyses were performed using Prism GraphPad software v9.0. statistical details of experiments can be found in the figure legends. All other statistical analyses were performed in RStudio, with associated packages where cited.

Nucleotide diversity analysis

We calculated nucleotide diversity, π, for each gene in the genome using two different datasets. Using the early isolate genomes in our dataset, we calculated π using the SNP calls we determined using Pilon. We used the –site-pi function in VCFtools to measure

nucleotide diversity on a per-site basis across all genomes.[38] For each gene in the reference genome, we calculated nucleotide diversity per gene by taking the average site diversity divided by the length of the gene.

To calculate π using a broader database of *S. enterica* Refseq complete genomes, we used the Refseq database construction tool from StrainGE to build a database,[73] starting with all 875 *S. enterica* Refseq complete genomes and clustering using a 90% Jaccard similarity threshold (which corresponds to 99.8% ANI) to remove highly similar genomes. This resulted in a database of 174 RefSeq genomes, representing the diversity of *S. enterica* [Table S5 Sheet 1]. In order to generate an alignment for each gene in our reference genome, *S.* Typhimurium SL1344, we first identified reciprocal best BLAST hits (RBHs) between each Refseq genome and our reference, and then used MUSCLE.[74] We calculated π for each gene alignment using MEGA X.[75]

Analysis of RNA-Seq data

Sequencing reads from each sample in a pool were demultiplexed based on their associated barcode sequence using custom scripts (https://github.com/broadinstitute/split_merge_pl). Up to 1 mismatch in the barcode was allowed, provided it did not make assignment of the read to a different barcode possible. Barcode sequences were removed from the first read, as were terminal G's from the second read, that may have been added by SMARTScribe during template switching.

Reads were aligned to *S.* Typhimurium *str.* SL1344, NC_016810.1, using BWA[77] and read counts were assigned to genes and other genomic features using custom scripts (https://github.com/broadinstitute/BactRNASeqCount). Read counts were used for differential expression analysis, conducted with edgeR (Version 3.32.1).[76] Differentially expressed genes were determined between the early and late isolates for each patient. Genes were determined as differentially expressed using a threshold of $p < 0.05$ and a false-discovery rate (FDR) of 5%, standard settings in EdgeR. EdgeR was used to calculate CPMs, Counts Per Million, for each gene, which were used for generating heatmaps with Pretty Heatmaps (version 1.0.12).