"Driver or passenger" : an integrated epidemiological and experimental perspective on the association between nontyphoidal salmonella infection and colon cancer

Duijster, J.W.

Chapter 8

FROM BEDSIDE TO BENCH: A MOLECULAR EPIDEMIOLOGICAL INVESTIGATION INTO THE ASSOCIATION BETWEEN NONTYPHOIDAL SALMONELLA INFECTION AND COLON CANCER

J.W. Duijster
V. Stévenin
C.E. Coipan
D.M. van Elsland A.H.A.M van Hoek
E.A. Neefjes-Borst
E. Franz
L. Mughini-Gras
J.J.C. Neefjes

In preparation

Abstract

Introduction: Epidemiological and experimental research suggests an association between nontyphoidal *Salmonella* (NTS) and colon cancer development. Yet, the drivers and potential mechanisms involved in this putative oncogenic role have not been deciphered. This study aims to unravel a causal link between NTS and colon cancer from the bacterial perspective.

Methods: We performed a matched case-control study based on NTS isolates obtained from 30 individuals who were diagnosed with colon cancer later in life (i.e., case-isolates) and 30 people without colon cancer diagnosis (i.e., control-isolates). All 60 NTS isolates were subjected to the following experiments/analyses: a) *in vitro* infection and host cell transformation assay; b) whole-genome sequencing; c) passage through an *in vitro* model system resembling the human gastrointestinal tract; d) *in vitro* quantification of different carbon (C-), nitrogen (N-), phosphorus (P-) and sulfur (S-) source utilization. The outcomes of the different experiments and analyses were used to assess whether case-isolates were different from control-isolates in terms of genotype or phenotype and whether this was associated with transformation efficiency.

Results: Substantial variation was present in the isolates' capacity to induce infection and cellular transformation *in vitro*, with a tendency towards higher transformation efficiency among the case-isolates. This could, however, not be explained by the genotype, neither were significant genotypic differences observed between case- and control-isolates. However, higher transformation efficiency was correlated with increased metabolic utilization capacity of multiple N-, P- and S-sources.

Conclusion: The outcomes of this study indicate a phenotypic rather than a genotypic driver for transformation efficiency. Yet, RNA sequencing of the isolates can reveal whether expression of genes differs between isolates with a high *versus* a low transformation efficiency.

Introduction

In the last decades, the role of bacteria in the onset and progression of cancers is being gradually acknowledged. Numerous mechanisms have been identified by which bacteria manipulate the host during infection, for instance by alteration of the host signaling pathways, the induction of chromosomal instability or prevention of apoptosis of damaged cells [1]. While *S*. Typhi and *Helicobacter pylori* as causative agents of respectively gallbladder and gastric cancer are well established, more species are added to the list of bacteria (potentially) contributing to cancer formation. An example hereof includes nontyphoidal *Salmonella* (NTS), which provokes cellular transformation in predisposed gallbladder organoids, as well as the development of colon tumors in mice after oral infection with the NTS serovar Typhimurium [2]. This was corroborated by an epidemiological study in the Netherlands where a significant positive association was found between infection with NTS and the risk of colon cancer (CC) [3]. In this registry-based study, routinely collected surveillance data of confirmed human NTS infections were linked at the person-level to nationwide CC diagnosis data. The overall risk of CC among people with a history of NTS infection between 20-60 years of age was 1.5-fold higher as compared to the general Dutch population. The association concerned particularly the proximal part of the colon, with over a two-fold increased risk, whereas no excess risk of distal CC was observed after NTS infection. The proximal colon is the most exposed part to NTS bacteria leaving the ileum (i.e., were NTS mainly resides). Moreover, the CC risk appeared to be higher after infection with the serovar Enteritidis as compared to Typhimurium or other serovars [3]. Whether the observed differences in risk estimates between the serovars can be explained by a difference in oncogenic capacity between NTS serovars or strains and which mechanisms and/or pathways are involved, is not yet known. Moreover, much about how transformation is maintained after *Salmonella* has been cleared from the host and which virulence factor(s) might be involved in the oncogenicity of NTS remains to be elucidated. To address these questions, we conducted a multi-faceted explorative study aiming to investigate the causal link between NTS infection and CC from bedside to bench, at a molecular epidemiology level. We first analyzed whether tumors obtained from people with a reported NTS infection are different in nature compared to tumors from people without such history. Second, we conducted a matched case-control study using a multi-angle approach, including *in vitro* and *in vivo* experiments to assess the infection, invasion, cellular transformation and metabolic capacities of NTS isolates as well as a genomic analysis of NTS isolates.

Materials and methods

Part A - Pathology

Formalin-fixed paraffin-embedded (FFPE) tissue blocks of 24 patients with proximal CC with a registered NTS infection in the past and 67 tissue blocks from age- and gender-matched controls (with proximal CC) without such history were obtained from the Dutch nationwide network and registry of histo- and cytopathology (PALGA).The tissue blocks were sectioned and the sections stained and counter-stained according to standard protocols. The sections were analyzed with immunohistochemistry for p53, c-MYC and MAPK/ERK. In addition, the tumor grading was determined for each of the 91 CC patients by an experienced CC pathologist.

Part B - Case-control study

We used data from the Dutch national surveillance program for *Salmonella* where public health laboratories send NTS isolates from human salmonellosis patients to the National Institute for Public Health and the Environment (RIVM) for serotyping [4]. This data was previously used in the epidemiological study in which 65 people with a reported NTS infection between January 1999-December 2015 were identified who developed proximal CC ≥1 year after the reported salmonellosis [3]. Of these 65 people, the NTS isolates from 30 people were still available for further analyses. We focused on cancers in the proximal part of the colon (ICD-10 codes C180-C185), as the association between NTS and CC was strongest for the proximal colon. Hence, we defined these 30 NTS isolates as cases. In addition, we selected NTS isolates (in the surveillance database) obtained from people with salmonellosis who did not develop CC during the period January 1999-December 2015 as controls. The 30 case isolates were matched on serovar, type of infection (enteric, septicemic, other [urinary, wound etc.]), year of infection, age at infection and gender to 30 controls (1:1 ratio), totaling 60 isolates.

Gastrointestinal tract model system

Prior to assessing the cellular transformation capacities of the 60 NTS isolates, we studied their host invasion and host infection potential in an *in vitro* model. The NTS isolates were cultured overnight at 37°C and subsequently exposed to conditions resembling the human digestive tract in a gastrointestinal tract (GIT) model system consisting of two parts: the simulated gastrointestinal passage and the attachment and invasion assay (Supplementary Figure S1) [5-7]. First, an overnight culture (ON) of each NTS isolate was sequentially exposed to simulated gastric fluid

(SGF) and simulated intestinal fluid (SIF) at 37°C for 30 minutes and 2 hours, respectively. After that, differentiated Caco-2 cells mimicking the small intestinal epithelium were inoculated with the SGF/SIF/bacterial-mixture at 37 °C for 1 hour on a 12-well plate to test the bacterial attachment (ATT) and invasion (INV) capacities. Between each step of the GIT model (ON, SGF, SIF, ATT and INV), serial 10-fold dilutions of samples were made and NTS bacteria present were enumerated. For quantification of attachment, 6 out of 12 wells containing the Caco-2 cells were lysed (to release attached and invaded bacteria), whereas for enumeration of invaded bacteria only, the other 6/12 wells were treated with gentamicin to kill attached bacteria before lysing cells to release invaded NTS. Details about the cell cultures, the dilutions steps, the compositions of SGF and SIF and the experimental procedures are described elsewhere [5-7]. The GIT model was applied for each of the 60 NTS isolates separately. For analysis of the change in bacterial count between each of the steps in the GIT model, we used a hierarchical Bayesian framework by applying Markov chain Monte Carlo sampling assuming Poisson distributed bacterial counts and lognormal distributed concentrations [6, 7]. Following the methodology of Wijnands et al. (2017), we calculated the *in vitro* infectivity (expressed as log P[inf]) as the sum of all log changes in NTS concentrations throughout the GIT model from the overnight culture until the concentration of invaded bacteria. The bacterial count data for all GIT system stages of the 60 NTS isolates were subjected to principal component analysis (PCA) to assess whether isolates obtained from cases differ from those obtained from controls in terms of their behavior/survival in the GIT model system [6]. Statistical analyses were performed in RStudio version 1.4 1103.

Bacterial strains and cell lines for NTS infection and transformation assays

S. Typhimurium strain SL1344 was a courtesy of S. Méresse [8]. This strain was used as reference strain in the *in vitro* infection and transformation assays. Mouse Embryonic Fibroblasts (MEFs) were derived from *Arf*-deficient C57BL/6 mice. MEFs overexpressing c-MYC were generated by retroviral transduction using a pLZRS-GFP(ires)-HA backbone. MEFs were cultured at 37°C, 5% CO2 in Dulbecco's Modified Eagle's Medium (DMEM) (Invitrogen) [2].

In vitro NTS infection, CFU and transformation assays

NTS infection of MEFs was performed as described previously [9]. In brief, NTS were grown overnight at 37°C in LB medium. The next day, the bacteria were sub-cultured at a dilution of 1:33 in fresh LB medium and incubated for 2 hours at 37°C while shaking. Cells were infected with NTS at multiplicity of infection (MOI) 20 in DMEM medium without antibiotics for 20 minutes at 37°C, 5% CO_2 in a tissue culture chamber and then incubated in the presence

of 100 µg/mL gentamicin (GIBCO) for 1 hour to eliminate extracellular bacteria. In the CFU assays, cells were lysed with lysis buffer (ddH2O + 1%NP-40), and serial dilutions of the lysate were plated on LB plates. In the anchorage-independent growth assays, MEFs were cultured for another 2 hours in the presence of 10 µg/mL gentamicin. The infected MEFs were subsequently collected and resuspended in DMEM medium supplemented with 10 µg/mL gentamicin and 0.35% low melting point agarose (UltraPure™, Invitrogen) and were poured on a soft agar bottom layer consisting of 0.7% low melting point agarose in DMEM with 10 µg/mL gentamicin. Anchorage-independent cell growth and number of soft agar colonies were assessed after 1-3 weeks of incubation at 37°C, 5% CO_2 using GelCountTM (Oxford Optronix, UK). For microscopy analysis, samples were fixed with 4% paraformaldehyde for 10 min at room temperature, and stained with rabbit polyclonal anti-*S*. Typhimurium LPS (Difco, Detroit, MI) and DAPI (Life Technologies). Images were acquired using a Leica TCS SP8 (Leica Microsystems, Wetzlar, Germany) at 40x or 63x magnification. Every experiment was performed in triplicate.

Genomic analysis

DNA extraction, whole genome sequencing (WGS) and assembly

As a next step in unraveling the role of NTS in cancer formation, we analyzed whether the degree of transformation capacity of the NTS isolates could be explained by differences in presence/absence of virulence genes. To this end, all 60 isolates were submitted to whole-genome sequencing (WGS). DNA isolation, 2×125 bp paired-end library preparation and WGS analysis on a HiSeq 2500 platform (Illumina) was performed by BaseClear (Leiden, the Netherlands). All resulting fastq files were subjected to quality control with CheckM v1.0.7 [10], and *de novo* assembled using an in-house developed pipeline (https://github.com/RIVM-bioinformatics/Juno_pipeline). The assembled genomes were analyzed with the SISTR application to confirm the *Salmonella* serovar [11]. Genome annotation of the assembled genomes was performed with Prokka v1.14.6 [12]. Next, the annotation data was used as input for Roary v3.13.0 to construct the core- and accessory genome of the 60 isolates, with a blastp identity cut-off of 95%, indicating a presence in at least 57/60 NTS isolates for a gene to be defined as part of the core-genome [13]. The core genome alignment from Roary was used to build a phylogenetic tree in RAxML v8.2.12 [14]. Single nucleotide polymorphisms (SNPs) were extracted using parsnp v1.2 [15]. The large number of serovars in the dataset, with consequently high genetic variability among the sequences, restrained us from calling SNPs on the full dataset. Instead, we created subsets comprising only *S*. Enteritidis, only *S*. Typhimurium, and both these serovars taken together. High density nucleotide

polymorphisms were filtered out with Gubbins v2.3.4 [16]. Protein function annotation was performed with Pannzer2 [17].

Genome-wide association analysis (GWAS)

Several genome-wide association tests were conducted using the R-package TreeWAS, which accounts for recombination and population structure [18]. Reconstruction of ancestral states was done with the parsimony method, and phylogenetic trees were constructed using maximum likelihood (ML) phylogeny. For all association tests, Bonferroni correction was applied to account for multiple testing. The outcome variable used in the association tests was the ranked transformation capacity of the bacterial strains, as inferred from the *in vitro* tests (see above). As differences in properties between bacterial strains may be the result of a modified structure of proteins or mutations in regulatory regions, which cannot be assessed from presence/absence of genes, we also performed the analysis at SNP (single nucleotide polymorphism) level. Statistical analyses were performed with R v3.6.2.

Phenotype Microarray analysis

In addition to the genomic analyses, we investigated phenotypic traits of the 60 NTS isolates by means of analyzing the utilization of carbon (C), nitrogen (N), phosphorus (P) and sulfur (S) sources. For this, we used the BioLog® Phenotype MicroArray (plates PM1, PM3 and PM4), which allows for high-throughput metabolic quantification of bacterial respiration and growth on a range of different substrates [19, 20]. Briefly, the quantification is based on redox technology, in which cell respiration is measured by the degree of irreversible reduction of a tetrazolium dye. Following Biolog instructions, a cell suspension of each individual cultured NTS isolate and a defined medium (including a dye) were added to 96-well plates containing a single C-, N-, P-, or S-source in each well. Plates were incubated at 37°C for 24h and color formation was measured every 15 minutes using an ELx808 Microplate Reader and Gen5 software (BioTek). The analyses were performed twice for each of the 60 NTS isolates. The ratio of the integrals of each C-, N-, P-, and S-source and a negative control (i.e. the PM1, PM3 and PM4 plates contain a negative control well without substrate for each source type), was used as outcome for further analysis. An hierarchical cluster analysis (HCA) and principal component analysis (PCA) was performed on the scaled data for each of the three plates to compare the metabolic phenotypes of NTS isolates obtained from cases *versus* controls taking into account the transformation capacity of the strains (as defined in the transformation assays). The analyses were performed in RStudio version 1.4 1103.

Results

Part A- Pathology

The objective of the pathology examination of the tumors was to determine whether the patients with a history of (severe) salmonellosis have different types of tumors with regard to the aforementioned markers as compared to the patients without such exposure. None of the markers (p53, c-MYC, MAPK/ERK) was significantly associated with a history of reported *Salmonella* infection. The tumors from patients with a history of *Salmonella* infection showed a tendency to be less likely undifferentiated than those of the patients without a reported *Salmonella* infection (odds ratio 0.21, 95% CI 0.04-1.06; p 0.059) (Figure 1, Supplementary Table S1 and Figure S2).

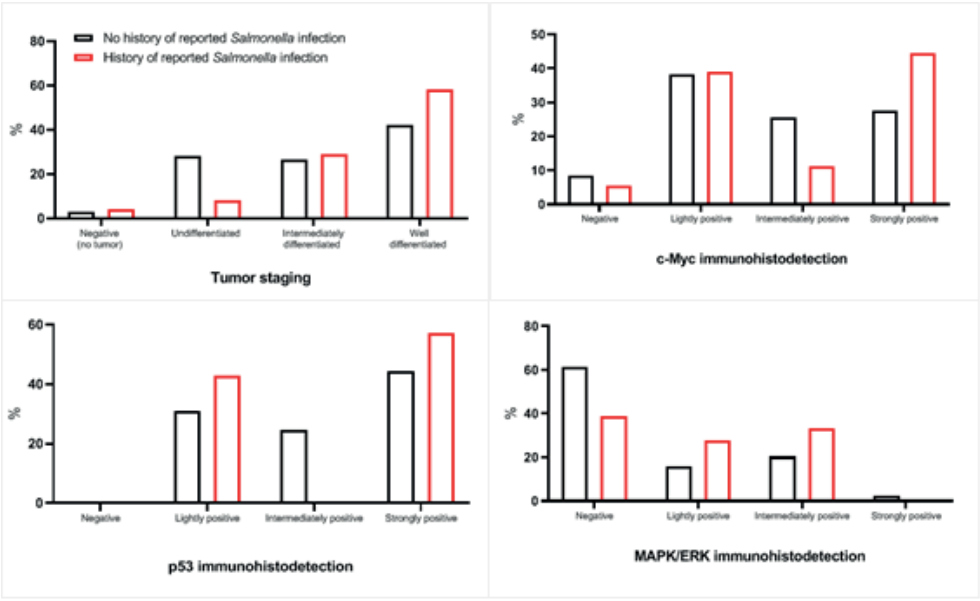


Figure 1. Immunohistochemistry and tumor staging results of the colon tumor blocks from patients with and without history of severe salmonellosis.

Part B – Case-control study

Description of study population

Supplementary table S2 shows the characteristics of the patients from whom the 30 case and 30 matched control NTS isolates were obtained. Two-third of the salmonellosis patients were male. The median age at infection was 63 years (interquartile range [IQR] 51-72) for controls and 67 years (IQR 55-76) for cases. Eleven different NTS serovars were included, mostly Enteritidis (n=22) and Typhimurium (including its monophasic variant) (n=18) (Figure 2, Supplementary table S2). Serovars other than Typhimurium and Enteritidis are hereafter referred to as 'other'. The vast majority (87%) of isolates were obtained from feces (i.e. enteric infections). One pair of Typhimurium isolates were obtained from blood or other normally sterile sites (i.e. invasive infections) and three pairs of isolates, belonging to the serovars Enteritidis, Typhimurium and Hadar, were obtained from urine or wound infections (Supplementary table S2).

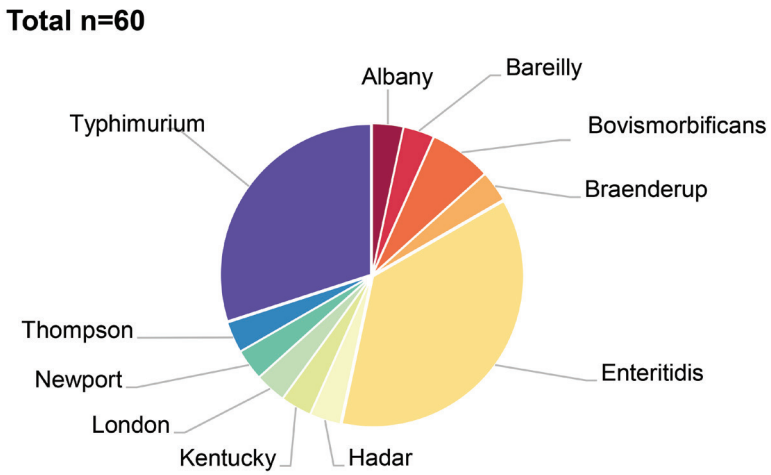


Figure 2. Serovar distribution of the 60 nontyphoidal *Salmonella* isolates.

Gastrointestinal tract model system

The mean *in vitro* infectivity (P[inf]) tends to be higher in NTS isolates obtained from cases (-1.74 ± 0.69; range -5.60 / -0.08) as compared to isolates obtained from controls (-1.39 ± 0.20; range -3.30 / -0.22), though (just) not statistically significant (paired t-test: t(29): 1.85, p 0.07). Figure 3 shows the log fractions of surviving bacteria in each of the four transitions

(SGF/ON, SIF/SGF, ATT/SIF, INV/ATT) of the GIT model for isolates obtained from cases *versus* controls as well as the mean *in vitro* infectivity. No statistically significant difference was observed between isolates obtained from cases *versus* controls for all four transitions (based on conditional logistic regression analysis). These results were confirmed by the outcomes of the PCA in which no clusters could be observed (Supplementary Figure 3). In fact, the data showed a negative correlation between the level of attachment (ATT_SIF) and invasion (INV_ATT), as well as between the amount of bacteria surviving the gastric fluid (SGF_ON) and those surviving intestinal fluid (SIF_SGF), as indicated by the arrows in Supplementary Figure 3.

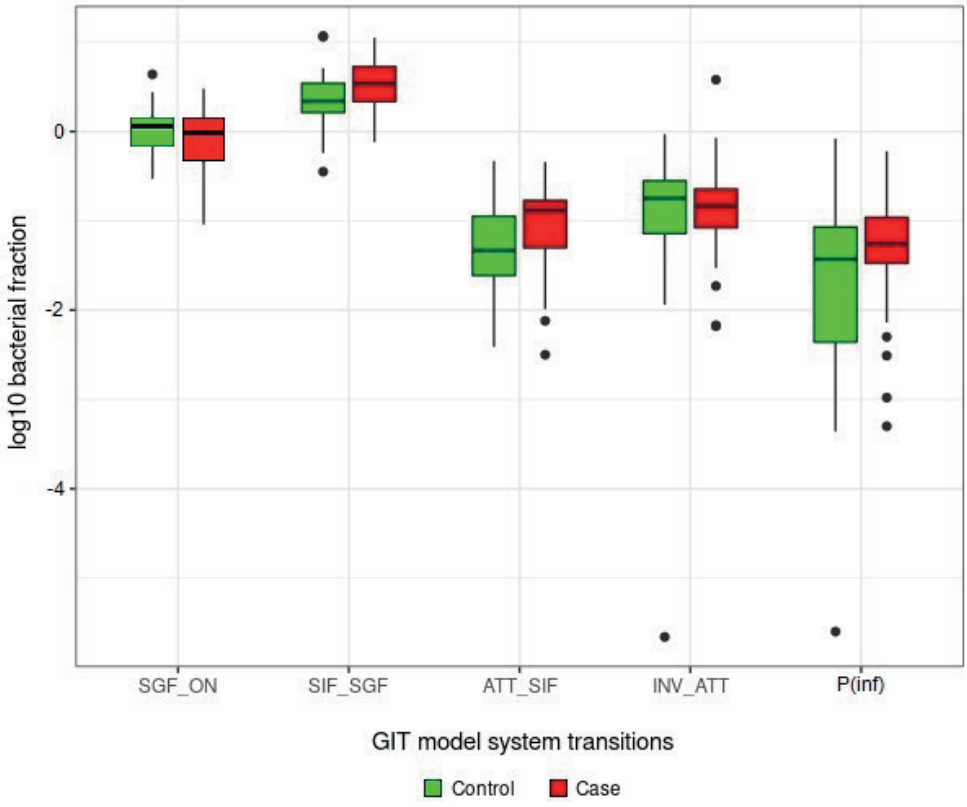


Figure 3. Log transformed fractions of bacterial counts in each of the transitions (survival of gastric fluid and intestinal fluid, attachment and invasion) in the gastrointestinal tract (GIT) model system for nontyphoidal *Salmonella* isolates obtained from cases vs. controls. SGF_ON: simulated gastric fluid vs. overnight bacterial culture; SIF_SGF: simulated intestinal fluid vs. simulated gastric fluid; ATT_SIF: attachment vs. simulated intestinal fluid; INV_ATT: invasion vs. attachment; P(inf): mean in vitro infectivity (INV/ON).

Infection and transformation assays

In the *in vitro* analyses, the infection and transformation capacity of the 60 NTS strains was assessed in *Arf-/-* + c-MYC MEFs. Infection and transformation capacities were normalized against the *S*. Typhimurium reference strain. Ten isolates, belonging to the serovars Enteritidis (n=4), Typhimurium (n=2), Albany (n=1), Bovismorbificans (n=1), Hadar (n=1) and Newport (n=1) failed to infect the MEFs (Figure 4). These included eight control isolates and two case isolates. Case isolates had a lower average infection efficiency compared to control isolates (0.55 *vs.* 0.58), whereas the opposite was true for the transformation efficiency (1.62 *vs.* 1.17) (Figure 4, Figure 5). Twenty-four isolates showed a higher infection efficiency as compared to the *S*. Typhimurium reference strain, 14 of these were case isolates (n=6 Typhimurium, n=4 Enteritidis, n=4 other serovars) (Figure 4, Supplementary Figure S4a, S4c). With regard to transformation efficiency, 18 case isolates (60%) and 10 control isolates (33%) showed a higher transformation efficiency as compared to the reference strain (Figure 5, Supplementary Figure S4b, S4d). Yet, the differences in infection and transformation capacity did not reach the level of significance (t-tests, both the infection and transformation capacities were not significantly higher for case isolates as compared to control isolates). Nonetheless, there was a tendency towards higher similar/higher infection efficiency in the case isolates (Figure 4). Likewise, case isolates were associated with higher transformation efficiencies, (Figure 5b, 5c).

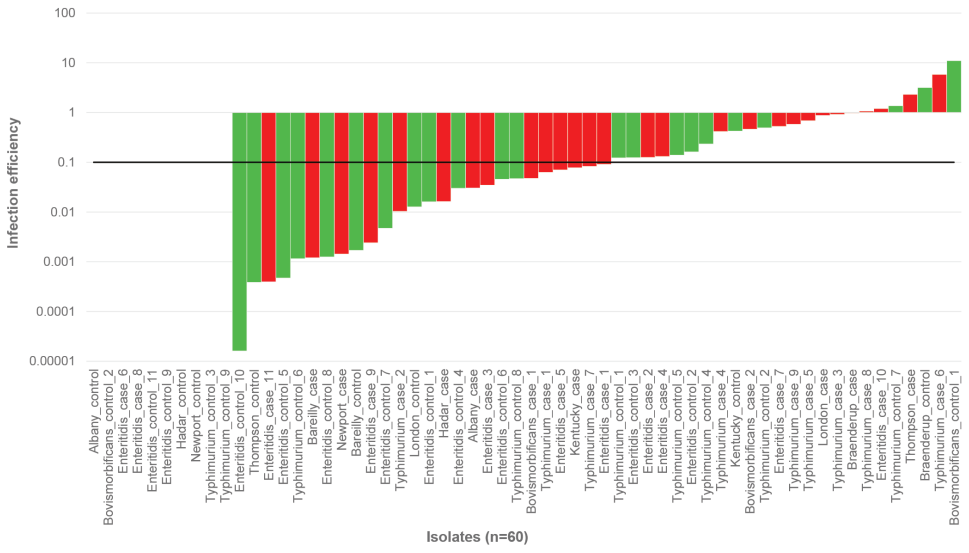


Figure 4. Mean infection potential of the 60 nontyphoidal *Salmonella* isolates obtained from cases (n=30; red bars) and controls (n=30; green bars) expressed in colony forming units (CFU) normalized against the infection efficiency of the *S*. Typhimurium reference strain. Results derive from three independent experiments with technical triplicates. Ten isolates were unable to infect mouse embryonic fibroblasts. Isolates with an infection efficiency >0.1 performed better than the laboratory strain.

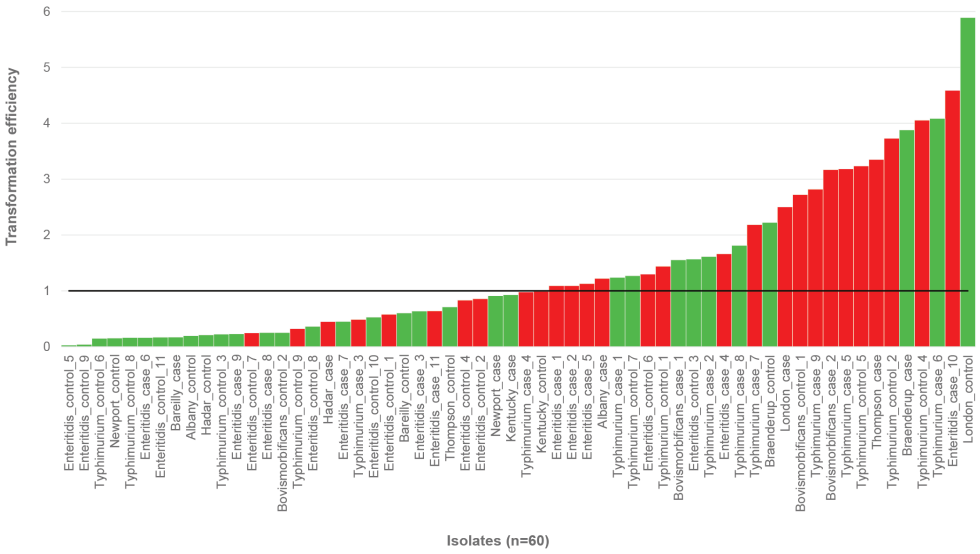


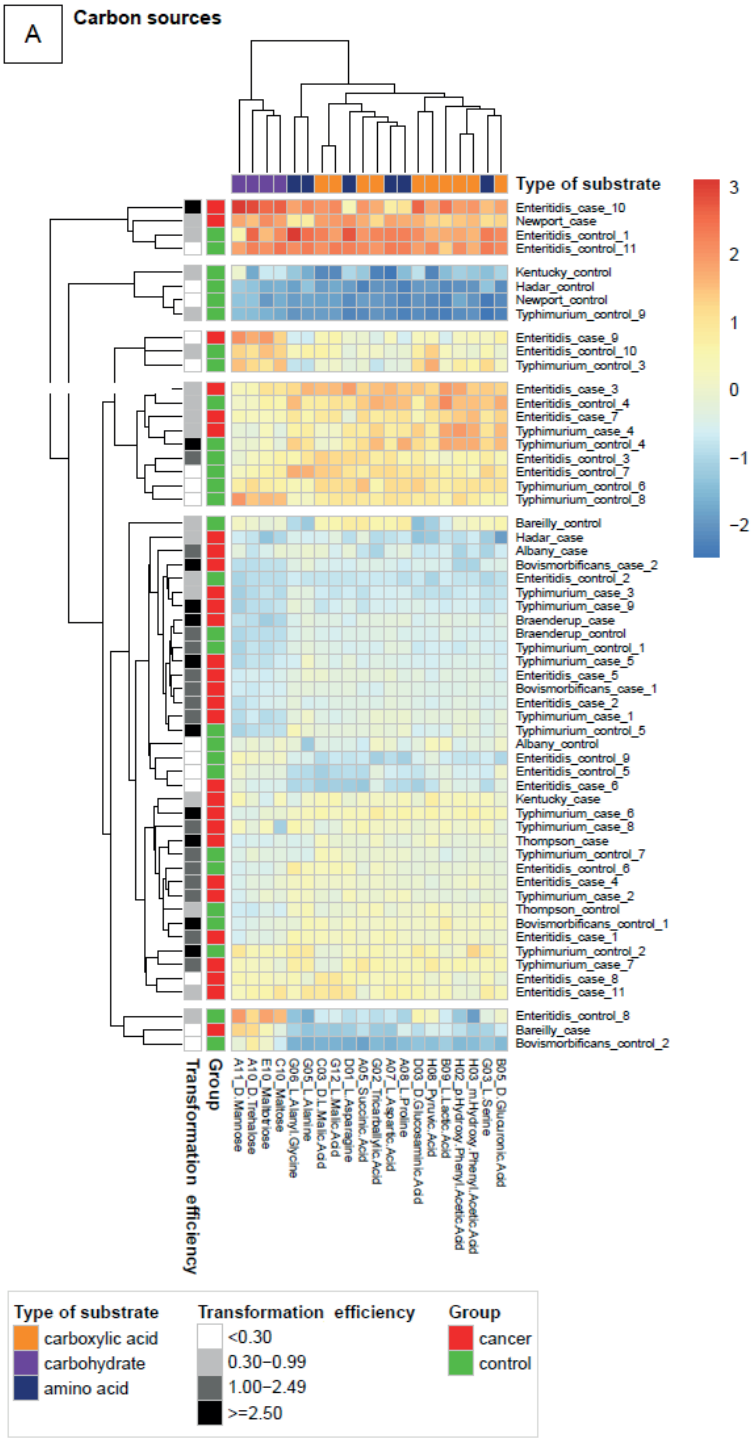
Figure 5. Mean transformation efficiencies of the nontyphoidal *Salmonella* isolates obtained from cases (n=30; red bars) and controls (n=30; green bars) normalized against the transformation efficiency of the *S*. Typhimurium reference strain. Results derive from three independent experiments with technical triplicates. Isolates with a transformation efficiency >1 performed better than the laboratory strain. Isolate 'London_control' shows an unexpected high transformation efficiency, likely as a result of sample contamination.

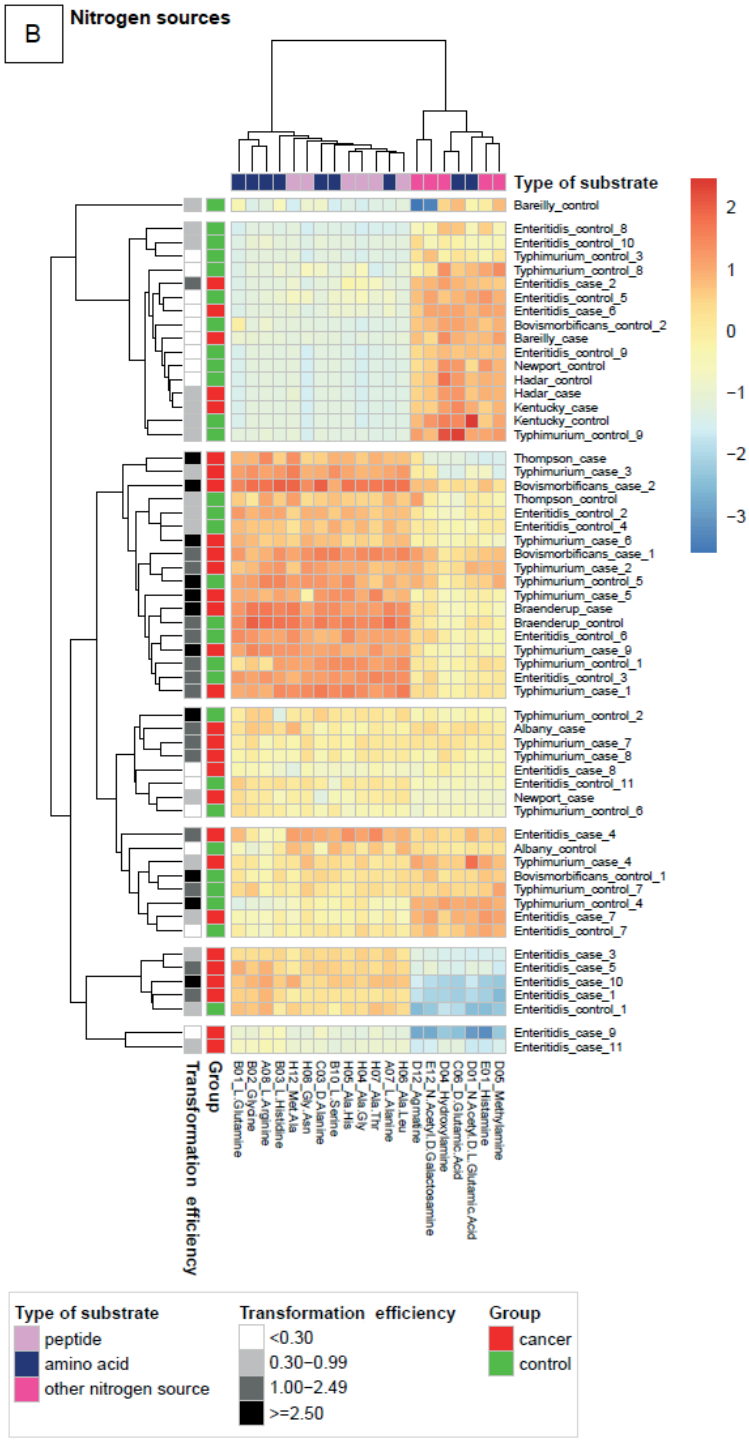
Genomic analysis

The genes in the pangenome of NTS as inferred by Prokka and Roary were used to inform the association tests performed by TreeWAS. We restricted the analysis to *S*. Enteritidis and *S*. Typhimurium isolates as these totaled 22 and 18 isolates respectively (in contrast to all other serovars with 2-4 isolates each). Presence of five genes appeared significantly associated with a higher transformation efficiency in the *S*. Typhimurium subset and four in the combined Enteritidis/Typhimurium subset (Supplementary Table S3). One of the significant genes of the *S*. Typhimurium subset and three from the combined Enteritidis/Typhimurium subset had unknown functions. For the remaining genes, functional annotation revealed three proteins involved in UV protection and mutation of which two are part of the bacterial SOS response to DNA damage (UmuC and UmuD). There were seven SNPs significantly associated with the ranked transformation efficiency of each of *S*. Enteritidis and *S*. Typhimurium subsets, and only one SNP significantly associated with the combined Enteritidis/Typhimurium subset. The genes in which these SNPs are located were associated with several functions including DNA cleavage and transcription activities (Supplementary Table S3).

Metabolic characterization of NTS isolates

In the metabolic analysis of bacterial phenotypes we analyzed 95 sole sources of carbon (PM1), 95 nitrogen sources (PM3) and 59 and 35 sole sources of respectively phosphorus and sulfur (PM4) (Supplementary Table S4). Conditional logistic regression was used to assess the association between substrate utilization and the likelihood of the host being diagnosed with colon cancer later in life. We observed a significant ($p<0.05$) positive association between isolates from cases and utilization of 11 substrates (10 phosphorus sources, 1 nitrogen source) and a negative association for 10 substrates (7 nitrogen sources, 3 carbon sources). Spearman's rank correlation test revealed 135 significant positive correlations and 34 significant negative correlations between substrate utilization and transformation efficiency (Supplementary Table S4). After Bonferroni correction for multiple testing, 76 positive correlations remained, comprising 34, 27 and 15 N- and P-/S-sources respectively. The remaining eight significant inverse correlations included 6 carbon sources and 2 sulfur sources (Supplementary Table S4). The tendency towards increased substrate utilization for isolates with a higher transformation efficiency was particularly pronounced for amino acids and several phosphorus sources. Before conducting a PCA, we defined the optimal clustering method for hierarchical clustering. The average linkage clustering showed the best fit to the data (lowest Gower distance and highest cophenetic correlation). Supplementary Figure S5 shows the heatmap of the scaled utilization scores of all 60 NTS isolates for the 76 positive correlations and eight inverse correlations with isolates being clustered using average linkage. PCAs on the scaled data revealed that the first two principal components (PCs) accounted for 58.7-63.5% of the total variance for the three plates (Supplementary Figures S6-S8). Ninety percent of the variance was explained by 13, 15 and 14 PCs for C-, N- and P-/S-sources respectively. Figure 6 shows the heatmaps of the nutrient utilization of the 60 isolates for the 20 nutrients with the highest contribution to the variance in the data, for the C-, N- and P-S-plate respectively. Particularly for several amino acids, peptides and phosphorus and sulfur sources the degree utilization tends to be higher for isolates with a higher transformation efficiency (part of the sources as also depicted in Supplementary Figure S5).


A
Carbon sources
Type of substrate
3
2
1
0
−1
−2
Enteritidis_case_10
Newport_case
Enteritidis_control_1
Enteritidis_control_11
Kentucky_control
Hadar_control
Newport_control
Typhimurium_control_9
Enteritidis_case_9
Enteritidis_control_10
Typhimurium_control_3
Enteritidis_case_3
Enteritidis_control_4
Enteritidis_case_7
Typhimurium_case_4
Typhimurium_control_4
Enteritidis_control_3
Enteritidis_control_7
Typhimurium_control_6
Typhimurium_control_8
Bareilly_control
Hadar_case
Albany_case
Bovismorbificans_case_2
Enteritidis_control_2
Typhimurium_case_3
Typhimurium_case_9
Braenderup_case
Braenderup_control
Typhimurium_control_1
Typhimurium_case_5
Enteritidis_case_5
Bovismorbificans_case_1
Enteritidis_case_2
Typhimurium_case_1
Typhimurium_control_5
Albany_control
Enteritidis_control_9
Enteritidis_control_5
Enteritidis_case_6
Kentucky_case
Typhimurium_case_6
Typhimurium_case_8
Thompson_case
Typhimurium_control_7
Enteritidis_control_6
Enteritidis_case_4
Typhimurium_case_2
Thompson_control
Bovismorbificans_control_1
Enteritidis_case_1
Typhimurium_control_2
Typhimurium_case_7
Enteritidis_case_8
Enteritidis_case_11
Enteritidis_control_8
Bareilly_case
Bovismorbificans_control_2
Transformation efficiency
Group
A11_D Mannose
A10_D Trehalose
E10_Maltotriose
C10_Maltose
G06_L Alanyl Glycine
G05_L Alanine
C03_D,L Malic Acid
G12_L Malic Acid
D01_L Asparagine
A05_Succinic Acid
G02_Tricarballylic Acid
A07_L Aspartic Acid
A08_L Proline
D03_D Glucosaminic Acid
H08_Pyruvic Acid
B09_L Lactic Acid
H02_p Hydroxy Phenyl Acetic Acid
H03_m Hydroxy Phenyl Acetic Acid
G03_L Serine
B05_D Glucuronic Acid
Type of substrate
carboxylic acid
carbohydrate
amino acid
Transformation efficiency
<0.30
0.30−0.99
1.00−2.49
>=2.50
Group
cancer
control


B
Nitrogen sources
Type of substrate
Bareilly_control
Enteritidis_control_8
Enteritidis_control_10
Typhimurium_control_3
Typhimurium_control_8
Enteritidis_case_2
Enteritidis_control_5
Enteritidis_case_6
Bovismorbificans_control_2
Bareilly_case
Enteritidis_control_9
Newport_control
Hadar_control
Hadar_case
Kentucky_case
Kentucky_control
Typhimurium_control_9
Thompson_case
Typhimurium_case_3
Bovismorbificans_case_2
Thompson_control
Enteritidis_control_2
Enteritidis_control_4
Typhimurium_case_6
Bovismorbificans_case_1
Typhimurium_case_2
Typhimurium_control_5
Typhimurium_case_5
Braenderup_case
Braenderup_control
Enteritidis_control_6
Typhimurium_case_9
Typhimurium_control_1
Enteritidis_control_3
Typhimurium_case_1
Typhimurium_control_2
Albany_case
Typhimurium_case_7
Typhimurium_case_8
Enteritidis_case_8
Enteritidis_control_11
Newport_case
Typhimurium_control_6
Enteritidis_case_4
Albany_control
Typhimurium_case_4
Bovismorbificans_control_1
Typhimurium_control_7
Typhimurium_control_4
Enteritidis_case_7
Enteritidis_control_7
Enteritidis_case_3
Enteritidis_case_5
Enteritidis_case_10
Enteritidis_case_1
Enteritidis_control_1
Enteritidis_case_9
Enteritidis_case_11
Transformation efficiency
Group
B01_L.Glutamine
B02_Glycine
A08_L.Arginine
B03_L.Histidine
H12_Met.Ala
H08_Gly.Asn
C03_D.Alanine
B10_L.Serine
H05_Ala.His
H04_Ala.Gly
H07_Ala.Thr
A07_L.Alanine
H06_Ala.Leu
D12_Agmatine
E12_N.Acetyl.D.Galactosamine
D04_Hydroxylamine
C06_D.Glutamic.Acid
D01_N.Acetyl.D.L.Glutamic.Acid
E01_Histamine
D05_Methylamine
2
1
0
−1
−2
−3
Type of substrate
peptide
amino acid
other nitrogen source
Transformation efficiency
<0.30
0.30−0.99
1.00−2.49
>=2.50
Group
cancer
control

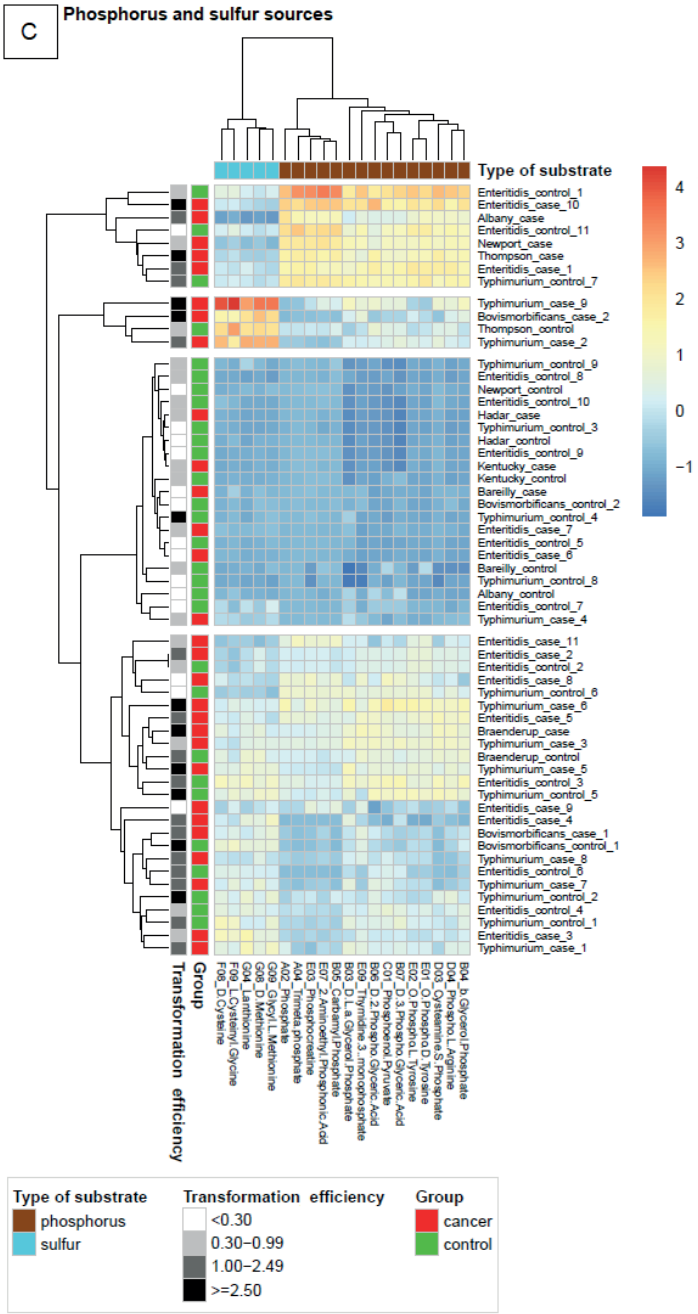


Figure 6. Heatmap of scaled utilization scores of the 60 nontyphoidal *Salmonella* (NTS) isolates for the top 20 sources mostly contributing to the variance in the carbon (A), nitrogen (B) and phosphorus/sulfur (C) data. NTS isolates are clustered based their utilization scores using the average linkage method. Left annotation depicts the group (case *vs.* control) and transformation efficiency (in four categories).

Discussion

Colon cancer ranks among the highest cancer incidences worldwide and the complexity of all microbial factors putatively contributing to the development of colon tumors is gradually being acknowledged. Literature on the role of microorganisms reveals a number of commensal and pathogenic bacteria associated with the induction of malignancies and progression of tumor growth. Yet, from a mechanistical perspective, a lot is unknown about possible pathways involved and whether the oncogenic potential of bacteria might be attributable to certain bacterial genes. For several bacteria, the oncogenic potential is restricted to strains/serotypes expressing specific genes or producing toxins [21-23]. With regard to NTS, multiple (effector) proteins have been identified that manipulate host cell-signaling pathways to escape immunity, reduce inflammation and apoptosis and enhance bacterial proliferation [1]. As collateral damage from host cell manipulation induced in the infection cycle, host cells can enter a cancerous state as part of the multistep process of cancer formation [1]. Whether the possible tumorigenic potential of NTS is attributable to serovar specific traits or genes is not yet investigated.

Here we describe the first study assessing possible genotypic or phenotypic traits of NTS related to cell transformation potential. Also we compared characteristics and markers of tumors derived from proximal colon cancer patients with a notified *Salmonella* infection in the past (*Salmonella*$^+$) to tumors from patients without such reported infection (*Salmonella*$^-$). A higher portion of tumors from *Salmonella*$^+$ patients were well-differentiated as compared to tumors from *Salmonella*$^-$ patients. Generally, well-differentiated (low-grade) tumors have a much better prognosis compared to poorly-differentiated tumors. These outcomes correspond to earlier research which revealed also a slightly higher percentage of well-differentiated proximal colon tumors in patients with a past *Salmonella* infection [3]. Immunohistochemistry of the tumor suppressor p53 and proto-oncogene c-MYC showed excess staining in tumors from *Salmonella*$^+$ patients, though not significant. Yet, many of the tissue samples were old and have been stored at different places throughout the Netherlands which presumably affected the labelling. Mutations in p53 and c-MYC are frequently observed in cancer patients, high levels of p53 staining are indicative for (mutated) inactivated p53 whereas c-MYC overexpression is presumably associated with tumor staging [2, 24-26]. Earlier experimental research showed that *S*. Typhimurium is able to induce tumor formation in mice as well as cellular transformation in gallbladder organoids and MEFs when these organisms/cell harbored inactivated p53 and overexpressed c-MYC [2]. The outcomes of a previous study corroborated this and we here demonstrate that also other NTS strains are able to induce the cellular transformations in MEFs [27].

Substantial variation in infection and transformation efficiency was observed between the isolates. Despite the lack of significant correlations between infectivity and transformation efficiency and disease outcome (cancer or no cancer), there was a tendency towards a higher transformation efficiency for the case-isolates. To our knowledge, this is the first study in which the transformation capacity was assessed for multiple NTS isolates allowing for comparison between isolates. In earlier studies, transformation assays were performed with *S*. Typhimurium [2, 27], yet our study clearly showed that other serovars might have an even higher transformation capacity. The results of the genomic analysis suggest several proteins involved in SOS response to be associated with transformation capacity in the *S*. Typhimurium subset. Nonetheless, these results should be interpreted with caution, as variants of the same gene might be differentially present between *S*. Typhimurium isolates. We observed this for the gene encoding the UmuC protein (data not shown). Whether a higher transformation capacity is related to restricted variants of genes (as could be for the UmuC gene) or the number of copies or variants present is not yet known. Furthermore, we observed several SNPs associated with transformation capacity in the *S*. Enteritidis, *S*. Typhimurium and combined Enteritidis/ Typhimurium subsets. The functions of the genes in which these SNPs were located were diverse and did not consistently indicate a possible role in cellular transformation.

The phenotypic microarray assay showed significant correlations between transformation efficiency and utilization scores of several nitrogen, phosphorus and sulfur sources. A high degree of metabolic flexibility confers a biological advantage for the bacterium as the availability and amount of nutrients changes during its infectious cycle [28]. Several putrefaction pathways have been identified by which commensal and pathogenic bacteria utilize amino acids released during fermentation of undigested proteins by resident bacteria [29]. During amino acid fermentation harmful metabolites such as H_2S, amines, phenol, indole and histidine are produced. Some of these metabolites have been associated with progression of colon cancer [29, 30]. How the metabolic signatures of specific NTS strains might induce a cascade of events and which pathways can be involved, will be the subject of further studies. A limitation of this study, which might have affected the observations, is the large portion (70%) of cases with a reported *Salmonella* infection above the age of 60. The risk of colon cancer increases substantially in older people, mostly as a result of the age-related accumulating of mutations associated with cancer [31]. This might have diluted some of the results, though unfortunately the sample size restrained us from restricting the analysis to the younger subgroups.

The outcomes of this study suggest that the oncogenic potential of *Salmonella* is better explained by phenotypic rather than genotypic traits of the isolates. Despite the absence of

relevant significant associations between gene presence and transformation in our study, we consider supplementing the genomic analyses with RNA sequencing analyses a good step forwards to improve our understanding of the underlying mechanisms. RNA sequencing provides information about the genes actually expressed rather than the presence/absence of genes and gene mutations as identified by DNA sequencing. This method revealed that the degree of pathogenicity among relatively genetically homogeneous strains of *S*. Enteritidis can be attributed to a multitude of genes differentially expressed between high and low pathogenic strains. Those genes were distributed over a range of functional classes including carbohydrate and amino acid metabolism, biogenesis and cell motility [32]. Applying RNA sequencing to the 60 NTS strains in our study might potentially identify differences between case- and control-isolates relevant for tumorigenesis.

Overall, this study revealed that tumors from colon cancer patients with a notified *Salmonella* infection in the past differ from tumors obtained from patients without such reported *Salmonella* infection with regard to tumor differentiation, yet tumor markers were not different between those groups. Moreover, the capacity to induce cellular transformation in MEFs varied between *Salmonella* isolates, with a tendency towards better transformation efficiency for isolates derived from people who were diagnosed with colon cancer later in life. This transformation efficiency was significantly correlated to utilization capacity of multiple nitrogen, phosphorus and sulfur sources. More in depth research is needed to unravel possible mechanisms and metabolic pathways which might be involved in the *Salmonella*-induced colon cancer development/progression.

References

1. Van Elsland D, Neefjes J. Bacterial infections and cancer. EMBO Reports. 2018;19(11):e46632.
2. Scanu T, Spaapen RM, Bakker JM, Pratap CB, Wu LE, Hofland I, et al. *Salmonella* manipulation of host signaling pathways provokes cellular transformation associated with gallbladder carcinoma. Cell Host & Microbe. 2015;17(6):763-74.
3. Mughini-Gras L, Schaapveld M, Kramers J, Mooij S, Neefjes-Borst EA, van Pelt W, et al. Increased colon cancer risk after severe *Salmonella* infection. PloS One. 2018;13(1):e0189721.
4. Van Pelt W, de Wit MA, Wannet WJ, Ligtvoet EJ, Widdowson MA, van Duynhoven YT. Laboratory surveillance of bacterial gastroenteric pathogens in The Netherlands, 1991-2001. Epidemiology and Infection. 2003;130(3):431-41.
5. Oliveira M, Wijnands L, Abadias M, Aarts H, Franz, E. Pathogenic potential of Salmonella typhimurium DT104 following sequential passage through soil, packaged fresh-cut lettuce and a model gastrointestinal tract. International Journal of Food Microbiology. 2011;148:149–55.
6. Wijnands LM, Teunis PF, Kuijpers AF, Asch DV, Ellen HM, Pielaat A. Quantification of Salmonella survival and infection in an in vitro model of the human intestinal tract as proxy for foodborne pathogens. Frontiers in Microbiology. 2017;8:1139.
7. Kuijpers AF, Bonacic Marinovic AA, Wijnands LM, Delfgou-van Asch EH, van Hoek AH, Franz E, et al. Phenotypic prediction: linking in vitro virulence to the genomics of 59 *Salmonella* enterica strains. Frontiers in Microbiology. 2019;9:3182.
8. Méresse S, Steele-Mortimer O, Finlay BB, Gorvel JP. The rab7 GTPase controls the maturation of *Salmonella* typhimurium-containing vacuoles in HeLa cells. The EMBO Journal. 1999;18:4394-4403.
9. Kuijl C, Savage NDL, Marsman M, Tuin AW, Janssen L, Egan DA, et al. Intracellular bacterial growth is controlled by a kinase network around PKB/AKT1. Nature. 2007; 450:725.
10. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research. 2015;25(7):1043-55.
11. Yoshida CE, Kruczkiewicz P, Laing CR, Lingohr EJ, Gannon VP, Nash JH, et al. The Salmonella in silico typing resource (SISTR): an open web-accessible tool for rapidly typing and subtyping draft *Salmonella* genome assemblies. PloS One. 2016;11(1):e0147101.
12. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics, 2014;30(14):2068-9.
13. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691-3.
14. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-3.
15. Treangen TJ, Ondov BD, Koren S, Phillippy, A. M. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biology. 2014;15(11):1-15.
16. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Research. 2015;43(3):e15-e15.
17. Collins C, Didelot X. A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. PLoS Computational Biology. 2018;14(2):e1005958.

18. Shea A, Wolcott M, Daefler S, Rozak DA. Biolog Phenotype Microarrays. In: Navid A. (eds) Microbial Systems Biology. Methods in Molecular Biology (Methods and Protocols), 2012;vol 881.
19. Biolog. (2021, November 23). Phenotype MicroArrays for Microbial Cells – Biolog. https://www.biolog.com/products-portfolio-overview/phenotype-microarrays-for-microbial-cells/
20. Hatakeyama M. Structure and function of Helicobacter pylori CagA, the first-identified bacterial protein involved in human cancer. Proceedings of the Japan Academy, Series B. 2017;93(4):196-219.
21. Cheng WT, Kantilal HK, Davamani, F. The mechanism of Bacteroides fragilis toxin contributes to colon cancer formation. The Malaysian journal of medical sciences: MJMS. 2020;27(4):9.
22. Pleguezuelos-Manzano C, Puschhof J, Huber AR, van Hoeck A, Wood HM, Nomburg J, et al. Mutational signature in colorectal cancer caused by genotoxic pks+ E. coli. Nature. 2020;580(7802):269-73.
23. Kaserer K, Schmaus J, Bethge U, Migschitz B, Fasching S, Walch A, et al. Staining patterns of p53 immunohistochemistry and their biological significance in colorectal cancer. The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland. 2000;190(4):450-6.
24. Toon CW, Chou A, Clarkson A, DeSilva K, Houang M, Chan JC, et al. Immunohistochemistry for myc predicts survival in colorectal cancer. PloS One. 2014;9(2):e87456.
25. Nakayama M, Sakai E, Echizen K, Yamada Y, Oshima H, Han TS, et al. Intestinal cancer progression by mutant p53 through the acquisition of invasiveness associated with complex glandular formation. Oncogene. 2017;36(42):5885-96.
26. Van Elsland DM, Duijster JW, Zhang J, Stévenin V, Zhang YG, Zha L, Xia Y, Franz E, Sun J, Mughini-Gras L, Neefjes JJC. Repetitive exposure to non-typhoidal Salmonellae is an environmental risk factor for colon cancer. Cell Host and Microbe 2021; submitted [see Chapter 4]
27. Reitzer L. Catabolism of amino acids and related compounds. EcoSal Plus. 2005;1(2).
28. Kaur H, Das C, Mande SS. In silico analysis of putrefaction pathways in bacteria and its implication in colorectal cancer. Frontiers in Microbiology. 2017;8:2166.
29. Gerner EW. Impact of dietary amino acids and polyamines on intestinal carcinogenesis and chemoprevention in mouse models. Biochemical Society Transactions. 2007;35(2):322-5.
30. Risques RA, Kennedy SR. Aging and the rise of somatic cancer-associated mutations in normal tissues. PLoS Genetics. 2018;14(1):e1007108.
31. Shah DH. RNA sequencing reveals differences between the global transcriptomes of *Salmonella* enterica serovar enteritidis strains with high and low pathogenicities. Applied and Environmental Microbiology, 2014;80(3):896-906.

Supplementary material

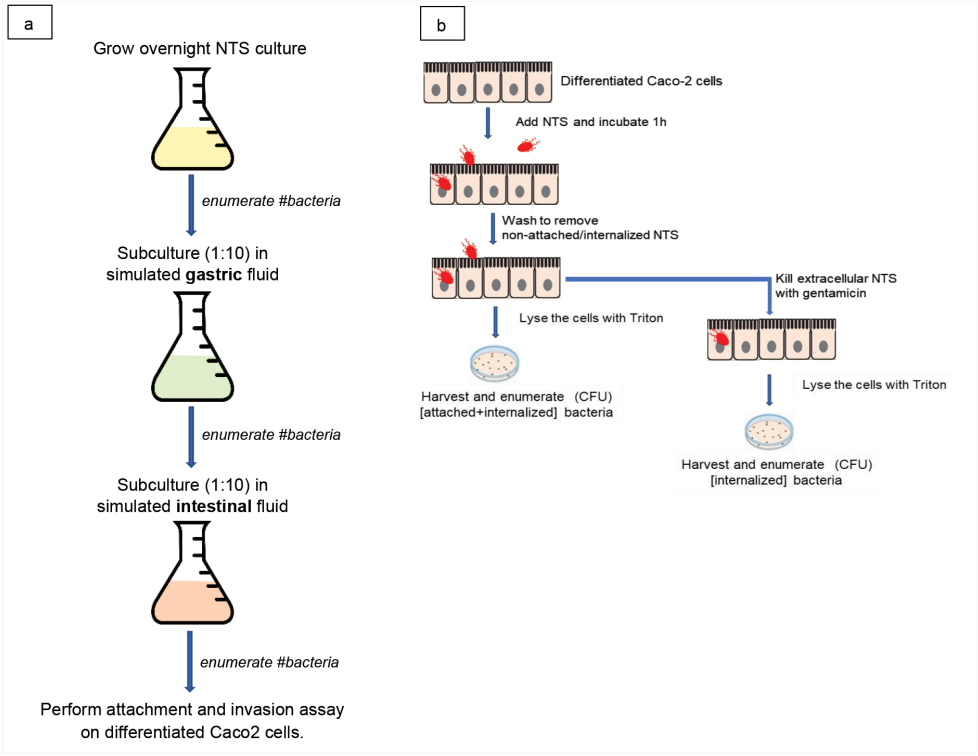


Figure S1. Schematic representation of the simulated gastrointestinal passage (A) and the attachment and invasion assay (B) of the gastrointestinal tract (GIT) model system. NTS: nontyphoidal *Salmonella*. Adapted from: Wijnands LM, Teunis PF, Kuijpers AF, Asch DV, Ellen HM, and Pielaat A. (2017). Quantification of *Salmonella* survival and infection in an in vitro model of the human intestinal tract as proxy for foodborne pathogens. Front. Microbiol., 8, 1139.

Table S1. Immunohistochemistry and tumor staging results of the colon tumor blocks from patients with and without history of severe salmonellosis.

	No history of reported NTS infection	History of reported NTS infection
Differentiation		
Negative (no tumor)	2 (3.13%)	1 (4.17%)
Well differentiated	27 (42.19%)	14 (58.33%)
Intermediately differentiated	17 (26.56%)	7 (29.17%)
Undifferentiated	18 (28.13%)	2 (8.33%)
c-MYC		
Negative	4 (8.51%)	1 (5.56%)
Lightly positive	18 (38.30%)	7 (38.89%)
Intermediately positive	12 (25.53%)	2 (11.11%)
Strongly positive	13 (27.66%)	8 (44.44%)
MAPK/ERK		
Negative	27 (61.36%)	7 (38.89%)
Lightly positive	7 (15.91%)	5 (27.78%)
Intermediately positive	9 (20.45%)	6 (33.33%)
Strongly positive	1 (2.27%)	0 (0.00%)
P53		
Negative	0 (0.00%)	0 (0.00%)
Lightly positive	19 (31.15%)	9 (42.86%)
Intermediately positive	15 (24.59%)	0 (0.00%)
Strongly positive	27 (44.26%)	12 (57.14%)

NTS: nontyphoidal *Salmonella.*

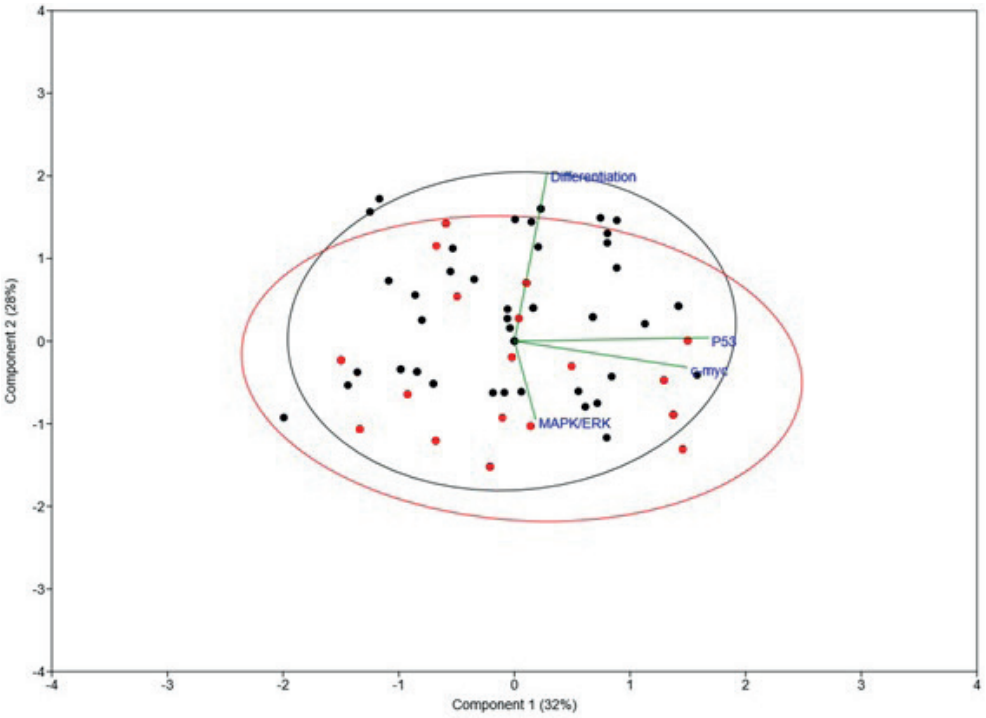


Figure S2. Principal component analysis of colon tumor blocks from colon cancer patients with (red dots) a without (black dots) reported history of *Salmonella* infection as a function of tumor differentiation and immunochemistry (markers p53, c-MYC, and fosfo-AKT/ERK). The principal components 1 and 2 respectively explain 32% and 28% of the total variance in the data.

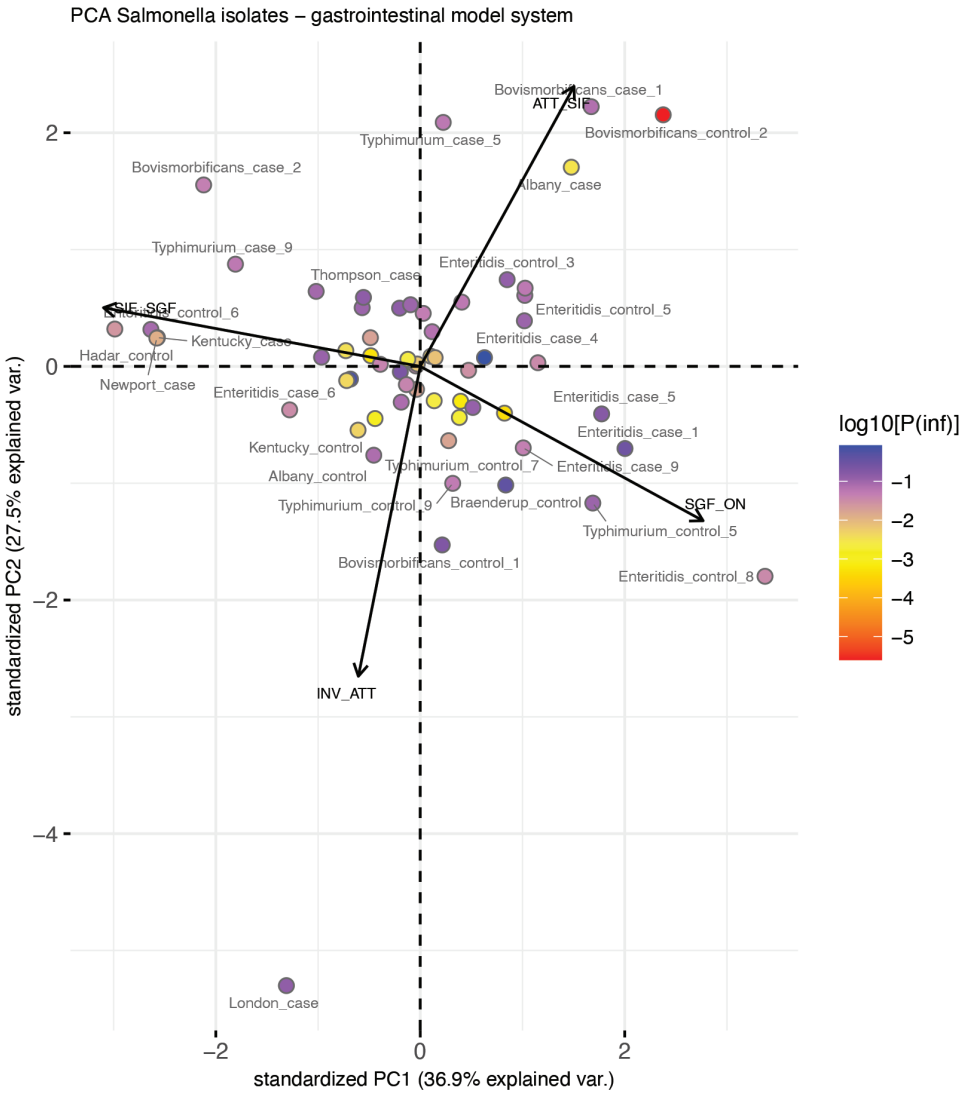


Figure S3. Principal component analysis (PCA) biplot of the fractions of enumerated bacteria in the steps of the gastrointestinal tract (GIT) model system for all 60 nontyphoidal *Salmonella* (NTS) isolates. The color scale of the dots reflects the log P(inf). The principal components (PC) 1 and PC2 explained 36.9% and 27.5% of the variance in the data respectively.

Table S2. Characteristics of the nontyphoidal *Salmonella* infection in individuals who developed colon cancer later in life (i.e. cases) versus those who did not develop cancer (i.e. controls).

Pair	Case/ control	Follow-up (years)	Serotype	Type of infection	Age at infection	Gender
#1	Case	13.6	Enteritidis	Enteric	50-54	Female
	Control	-	Enteritidis	Enteric	50-54	Female
#2	Case	10.4	Enteritidis	Enteric	35-39	Female
	Control	-	Enteritidis	Enteric	30-34	Female
#3	Case	5.4	Enteritidis	Enteric	70-74	Female
	Control	-	Enteritidis	Enteric	65-69	Female
#4	Case	2.6	Enteritidis	Enteric	40-44	Female
	Control	-	Enteritidis	Enteric	40-44	Female
#5	Case	9.9	Enteritidis	Other	65-69	Male
	Control	-	Enteritidis	Other	65-69	Male
#6	Case	1.3	Enteritidis	Enteric	45-49	Female
	Control	-	Enteritidis	Enteric	45-49	Female
#7	Case	3.4	Enteritidis	Enteric	55-59	Female
	Control	-	Enteritidis	Enteric	55-59	Female
#8	Case	7.8	Enteritidis	Enteric	75-79	Female
	Control	-	Enteritidis	Enteric	75-79	Female
#9	Case	4.0	Enteritidis	Enteric	75-79	Male
	Control	-	Enteritidis	Enteric	75-79	Male
#10	Case	4.0	Enteritidis	Enteric	75-79	Female
	Control	-	Enteritidis	Enteric	70-74	Female
#11	Case	1.9	Enteritidis	Enteric	70-74	Female
	Control	-	Enteritidis	Enteric	70-74	Female
#12	Case	5.4	Typhimurium	Enteric	60-64	Male
	Control	-	Typhimurium	Enteric	60-64	Male
#13	Case	13.7	Typhimurium	Enteric	35-39	Male
	Control	-	Typhimurium	Enteric	35-39	Male
#14	Case	3.4	Typhimurium	Enteric	65-69	Male
	Control	-	Typhimurium	Enteric	70-74	Male
#15	Case	2.1	Typhimurium*	Other	75-79	Male
	Control	-	Typhimurium*	Other	75-79	Female
#16	Case	1.1	Typhimurium	Enteric	50-54	Male
	Control	-	Typhimurium	Enteric	50-54	Male
#17	Case	4.3	Typhimurium	Septicemic	65-69	Male
	Control	-	Typhimurium*	Septicemic	60-64	Female
#18	Case	3.0	Typhimurium*	Enteric	70-74	Female
	Control	-	Typhimurium*	Enteric	70-74	Female
#19	Case	1.5	Typhimurium	Enteric	75-79	Female
	Control	-	Typhimurium*	Enteric	75-79	Female
#20	Case	1.1	Typhimurium*	Enteric	75-79	Female
	Control	-	Typhimurium*	Enteric	75-79	Female

Pair	Case/ control	Follow-up (years)	Serotype	Type of infection	Age at infection	Gender
#21	Case	3.1	Albany	Enteric	60-64	Female
	Control	-	Albany	Enteric	50-54	Female
#22	Case	4.3	Bareilly	Enteric	70-74	Male
	Control	-	Bareilly	Enteric	20-24	Female
#23	Case	7.1	Bovismorbificans	Enteric	65-69	Female
	Control	-	Bovismorbificans	Enteric	65-69	Female
#24	Case	8.0	Bovismorbificans	Enteric	60-64	Male
	Control	-	Bovismorbificans	Enteric	50-54	Female
#25	Case	11.5	Braenderup	Enteric	60-64	Female
	Control	-	Braenderup	Enteric	50-59	Male
#26	Case	1.1	Hadar	Other	75-79	Female
	Control	-	Hadar	Other	70-74	Female
#27	Case	1.9	Kentucky	Enteric	50-54	Male
	Control	-	Kentucky	Enteric	50-54	Male
#28	Case	4.2	London	Enteric	65-69	Female
	Control	-	London	Enteric	65-69	Female
#29	Case	4.3	Newport	Enteric	50-54	Female
	Control	-	Newport	Enteric	50-54	Female
#30	Case	2.5	Thompson	Enteric	75-79	Male
	Control	-	Thompson	Enteric	75-79	Male

*monophasic variant as shown by WGS analysis.

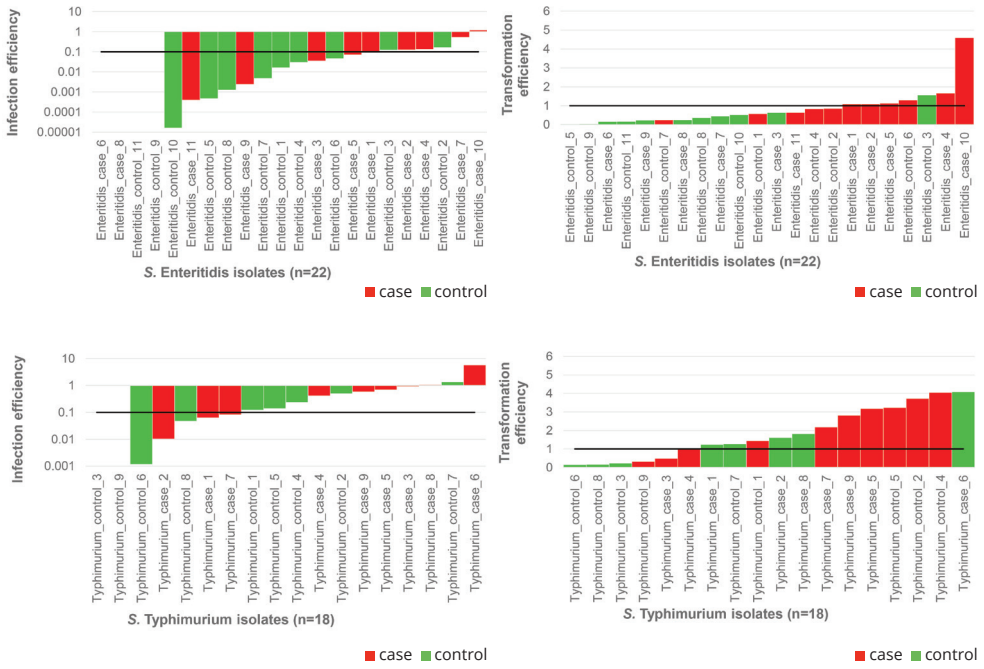


Figure S4. Mean infection and transformation efficiency of the 22 *S*. Enteritidis isolates (a, b) and 18 *S*. Typhimurium isolates (c, d) obtained from cases (red bars) and controls (green bars). Infection and transformation efficiencies are expressed in colony forming units normalized against the infection efficiency of the *S*. Typhimurium reference strain. Four Enteritidis isolates and two Typhimurium isolates were unable to infect mouse embryonic fibroblasts. Isolates with an infection efficiency of 0.1 and a transformation efficiency of 1.0 (black reference lines) performed better than the laboratory strain.

Table S3. Genes and SNPs significantly associated with transformation efficiency in *S*. Enteritidis, *S*. Typhimurium and combined Typhimurium/Enteritidis subsets.

Gene name	Annotation	Gene function
***S*. Typhimurium subset**		
group_2788	DNA-invertase hin [https://www.uniprot.org/uniprot/P03013]	Synthesis of phase-2 flagellin.
group_2807	protein ImpC [https://www.uniprot.org/uniprot/P0A1G0]	Belongs to the imp operon which has a function in UV protection and mutation.
group_1096	protein UmuC	Functions in UV protection and mutation, induced/SOS mutation.
group_420	*unknown*	*unknown*
umuD_2	protein umuD [https://www.uniprot.org/uniprot/P22493]	Functions in UV protection and mutation, induced/SOS mutation.
Combined *S*. Typhimurium and *S*. Enteritidis subset		
group_2065	*unknown*	*unknown*
group_2540	*unknown*	*unknown*
group_4752	*unknown*	*unknown*
group_4753	regulatory protein rop [https://www.uniprot.org/uniprot/P03051]	Regulatory role in plasmid DNA replication.
SNP locus name [gene name]	**Annotation**	**Gene function**
***S*. Enteritidis subset**		
502 [ehaB]	Autotransporter/virulence factor	Cell surface protein. Biofilm formation.
503 [res]	Type III restriction-modification system endonuclease	DNA cleavage.
504 [res]	Type III restriction-modification system endonuclease	DNA cleavage.
505 [res]	Type III restriction-modification system endonuclease	DNA cleavage.
506 [hsdR]	Type I restriction enzyme R protein	DNA cleavage. Nuclease and ATPase activities.
507 [*unknown*]	Cytoplasmic protein	*unknown*
508 [wcaI]	Glycosyltransferase, group 1 family protein	Glycosyltransferase activity.
***S*. Typhimurium subset**		
1926 [yegE]	Anti-FlhC(2)FlhD(4) factor YdiV	Transcription regulation. Virulence.
2058 [YacL]	UPF0231 protein YacL	*unknown*
2154 [*unknown*]	ISNCY family transposase	*unknown*
2367 [hxlA]	Putative hexulose 6 phosphate synthase	Formaldehyde fixation (ribulose monophosphate pathway).

Gene name	Annotation	Gene function
3604 [rfaB]	Lipopolysaccharide core heptose(I) kinase	Adding glycosyl residue to the core lipopolysaccharide. Detergent resistance.
3751 [rpoB]	DNA-directed RNA polymerase subunit beta	Catalyzation of the transcription of DNA into RNA.
4579 [*unknown*]	RpoE-regulated lipoprotein	*unknown*
Combined *S*. Typhimurium and *S*. Enteritidis subset		
20679 [oppA]	Periplasmic oligopeptide-binding protein	Peptide transmembrane transporter activity.

SNP: single nucleotide polymorphism.
IQR: interquartile range."

Table S4. Median and interquartile range of carbon, nitrogen, phosphorus and sulfur source utilization by case isolates (n=30) and control isolates (n=30) and Spearman correlation coefficient (rho) and p-value for the correlation between source utilization and transformation efficiency.

	Control isolates	Case isolates	Spearman		
	Median (IQR)	Median (IQR)	*rho*	p-value	Sig
PM1 – carbon sources					
L-Arabinose	1.72 (1.26-2.31)	1.62 (1.30-2.22)	-0.2350	0.071	
N-Acetyl-D-Glucosamine	2.44 (1.48-3.02)	2.06 (1.50-2.58)	-0.3716	<0.01	
D-Saccharic Acid	2.45 (1.01-3.56)	2.89 (1.27-3.22)	0.4127	<0.01	
Succinic Acid	3.03 (2.44-3.94)	3.01 (2.73-3.62)	0.1880	0.150	
D-Galactose	2.71 (1.88-3.56)	2.32 (2.03-3.22)	-0.3101	<0.05	
L-Aspartic Acid	2.91 (2.34-3.71)	3.03 (2.62-3.47)	0.1431	0.276	
L-Proline	3.13 (2.35-3.93)	3.10 (2.76-3.79)	0.2184	0.094	
D-Alanine	2.58 (1.60-3.08)	2.58 (2.24-3.02)	0.3207	<0.05	
D-Trehalose	2.79 (1.82-3.83)	2.41 (1.88-3.20)	-0.2954	<0.05	
D-Mannose	2.35 (1.73-3.04)	2.04 (1.72-2.84)	-0.3079	<0.05	
Dulcitol	2.62 (2.04-3.45)	2.49 (1.98-3.16)	-0.2152	0.099	
D-Serine	3.52 (2.68-4.25)	3.32 (2.97-3.60)	0.0699	0.595	
D-Sorbitol	2.97 (2.45-3.47)	2.78 (2.20-3.12)	-0.1900	0.146	
Glycerol	3.31 (2.56-4.32)	3.24 (2.88-3.74)	0.1173	0.372	
L-Fucose	2.37 (2.04-3.14)	2.25 (1.94-3.04)	-0.1807	0.167	
D-Glucuronic Acid	3.95 (3.07-4.58)	3.70 (3.28-4.47)	0.0440	0.739	
D-Gluconic Acid	3.88 (3.28-4.79)	3.97 (3.55-4.79)	0.1120	0.394	
D,L-α-Glycerol-Phosphate	3.49 (2.66-4.06)	3.31 (2.93-3.86)	0.1332	0.310	
D-Xylose	2.03 (1.36-2.83)	2.41 (1.68-3.05)	0.1695	0.195	
L-Lactic Acid	3.08 (2.56-3.58)	3.03 (2.72-3.61)	0.1313	0.317	
Formic Acid	0.98 (0.93-1.13)	0.99 (0.92-1.10)	0.3365	<0.05	
D-Mannitol	2.50 (1.74-3.20)	1.97 (1.65-2.74)	-0.3204	<0.05	
L-Glutamic Acid	2.18 (1.85-2.88)	2.00 (1.75-2.44)	-0.0356	0.787	
D-Glucose-6-Phosphate	4.19 (3.72-4.89)	4.38 (3.65-5.09)	-0.0166	0.900	
D-Galactonic Acid-y-Lactone	1.92 (1.29-2.50)	1.45 (1.06-1.91)	-0.5263	<0.001	+

	Control isolates	Case isolates	Spearman		
	Median (IQR)	Median (IQR)	*rho*	p-value	Sig
D,L-Malic Acid	2.77 (2.14-3.60)	2.86 (2.53-3.44)	0.1081	0.411	
D-Ribose	2.46 (1.48-3.19)	2.44 (2.01-2.84)	0.0543	0.680	
Tween 20	2.24 (1.72-2.95)	2.19 (1.95-2.84)	0.2136	0.101	
L-Rhamnose	2.06 (1.54-2.54)	1.66 (1.43-2.20)	-0.3300	<0.05	
D-Fructose	2.83 (1.88-3.26)	2.31 (1.84-2.88)	-0.3292	<0.05	
Acetic Acid	2.13 (1.37-2.52)	2.11 (1.83-2.60)	0.1909	0.144	
α-D-Glucose	2.57 (1.93-3.16)	2.02 (1.62-2.83)	-0.2838	<0.05	
Maltose	2.74 (2.13-3.76)	2.63 (2.19-3.07)	-0.2182	0.094	
D-Melibiose	2.76 (2.16-4.35)	2.87 (2.07-3.39)	-0.2008	0.124	
Thymidine	3.34 (2.69-3.78)	3.15 (2.66-3.56)	0.1426	0.277	
L-Asparagine	2.87 (2.28-3.77)	2.93 (2.64-3.42)	0.2035	0.119	
D-Aspartic Acid	1.91 (1.56-2.31)	2.01 (1.80-2.33)	0.2859	<0.05	
D-Glucosaminic Acid	3.09 (2.43-4.05)	3.08 (2.67-3.63)	0.1207	0.358	
1,2-Propanediol	0.98 (0.91-1.05)	0.99 (0.86-1.04)	-0.0462	0.726	
Tween 40	2.48 (1.72-3.17)	2.41 (2.25-2.86)	0.3645	<0.01	
α-Keto-Glutaric Acid	1.18 (1.11-1.34)	1.23 (1.12-1.33)	0.3115	<0.05	
α-Keto-Butyric Acid	1.96 (1.44-2.57)	2.02 (1.80-2.31)	0.0471	0.721	
α-Methyl-D-Galactoside	3.13 (2.49-4.22)	2.74 (2.31-3.52)	-0.1918	0.142	
α-D-Lactose	1.18 (1.06-1.32)	1.23 (1.12-1.28)	0.3368	<0.05	
Lactulose	0.86 (0.63-0.97)	0.70 (0.57-0.85)	-0.4206	<0.01	
Sucrose	1.05 (1.01-1.14)	1.13 (0.98-1.22)	0.1119	0.395	
Uridine	3.95 (2.76-5.17)	3.97 (3.48-4.96)	0.1195	0.363	
L-Glutamine	1.95 (1.54-2.21)	1.85 (1.59-2.34)	-0.0308	0.815	
M-Tartaric Acid	2.07 (1.23-2.83)	2.12 (1.55-2.34)	0.0718	0.585	
D-Glucose-1-Phosphate	3.06 (1.90-3.90)	3.41 (2.34-3.90)	0.2860	<0.05	
D-Fructose-6-Phosphate	3.78 (2.66-4.51)	3.82 (3.43-4.64)	0.3266	<0.05	
Tween 80	1.98 (1.47-2.45)	2.04 (1.87-2.45)	0.3719	<0.01	
α-Hydroxy Glutaric Acid-γ-Lactone	1.13 (0.92-1.31)	1.03 (0.78-1.18)	-0.2633	<0.05	
α-Hydroxy Butyric Acid	1.88 (1.56-2.64)	2.05 (1.83-2.28)	0.1966	0.132	
ß-Methyl-D-Glucoside	1.45 (1.29-1.76)	1.66 (1.33-1.80)	0.2906	<0.05	
Adonitol	1.10 (1.05-1.18)	1.10 (1.01-1.16)	-0.0123	0.926	
Maltotriose	3.01 (2.48-4.01)	2.79 (2.48-3.59)	-0.2102	0.107	
2-Deoxy Adenosine	4.01 (3.39-5.13)	3.81 (3.33-4.54)	-0.1582	0.227	
Adenosine	4.37 (3.40-5.05)	3.84 (3.38-4.57)	-0.127	0.335	
Glycyl-L-Aspartic Acid	2.47 (1.98-2.95)	2.18 (2.05-2.75)	0.138	0.294	
Citric Acid	2.95 (1.37-3.63)	2.99 (2.68-3.44)	0.359	<0.01	
M-Inositol	0.99 (0.89-1.21)	0.90 (0.72-1.52)	0.010	0.940	
D-Threonine	0.95 (0.88-1.16)	0.93 (0.80-1.05)	-0.268	<0.05	
Fumaric Acid	2.88 (2.26-3.37)	2.85 (2.37-3.20)	0.015	0.913	
Bromo Succinic Acid	2.35 (1.71-2.98)	2.35 (1.96-2.80)	0.155	0.237	
Propionic Acid	2.29 (1.30-2.92)	2.35 (2.04-2.71)	0.278	<0.05	
Mucic Acid	2.11 (1.05-3.45)	2.52 (1.11-3.11)	0.298	<0.05	
Glycolic Acid	0.94 (0.80-0.99)	0.86 (0.75-1.01)	-0.135	0.303	

	Control isolates	Case isolates	Spearman		
	Median (IQR)	Median (IQR)	*rho*	p-value	Sig
Glyoxylic Acid	1.37 (0.98-1.76)	1.39 (1.07-1.66)	0.223	0.087	
D-Cellobiose	1.29 (1.16-1.86)	1.28 (1.19-1.90)	-0.044	0.739	
Inosine	4.26 (3.37-5.42)	3.95 (3.33-5.46)	-0.108	0.411	
Glycyl-L-Glutamic Acid	2.45 (1.76-2.87)	2.28 (2.20-2.83)	0.137	0.296	
Tricarballylic Acid	3.41 (2.48-4.07)	3.39 (2.85-3.85)	0.204	0.118	
L-Serine	3.40 (2.84-4.32)	3.44 (2.96-3.97)	0.102	0.436	
L-Threonine	1.34 (1.02-1.76)	1.21 (1.04-1.77)	-0.118	0.370	
L-Alanine	2.44 (1.56-3.08)	2.63 (2.30-3.00)	0.345	<0.05	
L-Alanyl-Glycine	2.67 (2.02-3.32)	2.64 (2.41-3.36)	0.283	<0.05	
Acetoacetic Acid	1.00 (0.84-1.19)	0.99 (0.88-1.14)	-0.125	0.340	
N-Acetyl-ß-D-Mannosamine	3.27 (2.73-3.91)	3.01 (2.72-3.47)	0.056	0.672	
Mono Methyl Succinate	1.10 (0.93-1.33)	1.09 (0.86-1.36)	-0.390	<0.01	
Methyl Pyruvate	3.21 (2.67-4.01)	2.89 (2.53-3.73)	-0.171	0.192	
D-Malic Acid	0.81 (0.63-0.96)	0.67 (0.55-0.86)	-0.523	<0.001	+
L-Malic Acid	2.87 (2.24-3.56)	2.71 (2.46-3.44)	0.076	0.562	
Glycyl-L-Proline	2.60 (1.88-3.09)	2.56 (2.24-2.91)	0.314	<0.05	
p-Hydroxy Phenyl Acetic Acid	3.02 (2.31-3.77)	2.92 (2.64-3.47)	0.146	0.266	
m-Hydroxy Phenyl Acetic Acid	3.01 (2.36-3.73)	2.98 (2.65-3.51)	0.182	0.164	
Tyramine	2.76 (1.31-3.58)	2.95 (2.62-3.35)	0.375	<0.01	
D-Psicose	1.51 (1.27-1.80)	1.47 (1.31-1.79)	0.028	0.829	
L-Lyxose	0.78 (0.70-0.98)	0.75 (0.62-0.88)	-0.347	<0.05	
Glucuronamide	1.01 (0.76-1.10)	0.82 (0.68-0.97)	-0.518	<0.001	+
Pyruvic Acid	3.64 (2.85-4.59)	3.46 (2.95-4.31)	0.025	0.848	
L-Galactonic Acid-γ-Lactone	0.92 (0.69-1.01)	0.72 (0.65-0.94)	-0.485	<0.001	+
D-Galacturonic Acid	0.99 (0.79-1.07)	0.85 (0.75-1.03)	-0.427	<0.01	
Phenylethylamine	0.77 (0.64-0.95)	0.69 (0.59-0.81)	-0.482	<0.001	+
2-Aminoethanol	0.84 (0.64-0.98)	0.71 (0.56-0.89)	-0.534	<0.001	+
PM3 – Nitrogen sources					
Ammonia	1.00 (0.98-1.12)	1.00 (0.98-1.72)	-0.044	0.738	
Nitrite	0.88 (0.85-0.96)	0.87 (0.80-0.92)	-0.367	<0.01	
Nitrate	0.96 (0.93-0.99)	0.94 (0.87-0.99)	-0.072	0.584	
Urea	0.95 (0.92-1.00)	0.93 (0.85-0.98)	-0.126	0.337	
Biuret	0.98 (0.93-1.02)	0.95 (0.85-1.00)	-0.023	0.862	
L-Alanine	2.77 (1.02-3.91)	3.43 (2.13-3.84)	0.639	<0.001	+
L-Arginine	1.83 (1.10-3.40)	2.94 (1.72-3.58)	0.649	<0.001	+
L-Asparagine	2.54 (1.01-4.26)	3.71 (2.04-4.18)	0.607	<0.001	+
L-Aspartic Acid	1.47 (1.01-3.77)	2.74 (1.65-3.96)	0.570	<0.001	+
L-Cysteine	6.56 (1.07-8.18)	6.69 (3.09-8.41)	0.541	<0.001	+
L-Glutamic Acid	1.94 (1.12-3.03)	2.36 (1.58-3.20)	0.467	<0.001	+
L-Glutamine	3.19 (1.17-4.34)	3.97 (2.73-4.72)	0.632	<0.001	+
Glycine	2.20 (1.04-3.56)	3.31 (1.97-3.87)	0.682	<0.001	+
L-Histidine	2.83 (1.09-3.92)	3.43 (2.07-4.46)	0.595	<0.001	+
L-Isoleucine	0.99 (0.97-1.01)	1.00 (0.91-1.04)	0.187	0.152	

	Control isolates	Case isolates	Spearman		
	Median (IQR)	Median (IQR)	*rho*	p-value	Sig
L-Leucine	0.99 (0.96-1.07)	1.01 (0.91-1.10)	0.100	0.447	
L-Lysine	1.05 (1.01-1.14)	1.14 (1.01-1.44)	0.657	<0.001	+
L-Methionine	1.01 (0.97-1.07)	1.02 (0.98-1.12)	0.430	<0.01	
L-Phenylalanine	1.08 (1.00-1.47)	1.25 (1.05-1.61)	0.428	<0.01	
L-Proline	4.23 (1.22-5.30)	4.40 (3.04-5.66)	0.648	<0.001	+
L-Serine	2.98 (1.00-4.52)	3.91 (2.02-4.78)	0.646	<0.001	+
L-Threonine	1.23 (1.04-1.98)	1.40 (1.09-2.17)	0.527	<0.001	+
L-Tryptophan	1.01 (0.95-1.05)	0.98 (0.90-1.03)	-0.103	0.434	
L-Tyrosine	1.08 (1.04-1.14)	1.03 (0.96-1.09)	-0.118	0.369	
L-Valine	1.00 (0.98-1.03)	0.98 (0.92-1.02)	0.204	0.118	
D-Alanine	2.84 (1.22-3.91)	3.66 (1.80-4.54)	0.689	<0.001	+
D-Asparagine	1.05 (1.00-1.18)	1.08 (1.02-1.58)	0.323	<0.05	
D-Aspartic Acid	1.08 (1.04-1.36)	1.09 (1.04-1.33)	0.342	<0.05	
D-Glutamic Acid	0.86 (0.79-0.93)	0.81 (0.70-0.89)	-0.332	<0.05	
D-Lysine	1.09 (1.02-1.15)	1.07 (1.00-1.14)	0.323	<0.05	
D-Serine	4.07 (1.51-4.71)	4.21 (3.52-5.08)	0.657	<0.001	+
D-Valine	1.09 (1.01-1.16)	1.16 (1.00-1.33)	0.508	<0.001	+
L-Citrulline	1.03 (0.99-1.14)	1.08 (0.99-1.26)	0.520	<0.001	+
L-Homoserine	0.85 (0.80-0.94)	0.83 (0.74-0.92)	-0.187	0.160	
L-Ornithine	1.11 (1.00-1.18)	1.14 (1.02-1.42)	0.435	<0.01	
N-Acetyl-D,L-Glutamic Acid	0.95 (0.92-1.01)	0.93 (0.85-1.01)	-0.139	0.289	
N-Phthaloyl-LGlutamic Acid	1.03 (0.94-1.12)	0.98 (0.85-1.06)	-0.013	0.920	
L-Pyroglutamic Acid	0.97 (0.93-1.00)	0.96 (0.86-1.01)	-0.132	0.317	
Hydroxylamine	0.80 (0.75-0.90)	0.79 (0.66-0.85)	-0.325	<0.05	
Methylamine	0.96 (0.91-1.02)	0.91 (0.81-1.00)	-0.271	<0.05	
N-Amylamine	1.00 (0.96-1.03)	0.98 (0.88-1.05)	-0.088	0.503	
N-Butylamine	1.03 (1.00-1.06)	1.04 (0.93-1.06)	-0.033	0.804	
Ethylamine	1.04 (1.02-1.09)	1.03 (0.97-1.09)	0.153	0.242	
Ethanolamine	1.03 (0.99-1.04)	1.01 (0.93-1.07)	0.118	0.368	
Ethylenediamine	0.98 (0.93-1.07)	0.96 (0.85-1.03)	-0.115	0.381	
Putrescine	1.01 (0.98-1.05)	1.01 (0.92-1.05)	0.140	0.285	
Agmatine	1.00 (0.96-1.03)	0.99 (0.89-1.03)	-0.023	0.863	
Histamine	1.04 (0.98-1.08)	0.99 (0.91-1.08)	-0.237	0.069	
ß-Phenylethyl-amine	0.97 (0.94-1.03)	0.96 (0.90-1.01)	-0.094	0.476	
Tyramine	3.34 (1.27-4.21)	3.76 (2.60-4.54)	0.653	<0.001	+
Acetamide	1.06 (1.03-1.11)	1.05 (1.01-1.09)	0.141	0.282	
Formamide	1.07 (1.03-1.10)	1.04 (1.01-1.09)	0.147	0.261	
Glucuronamide	1.71 (1.26-2.10)	1.66 (1.46-1.92)	0.333	<0.05	
D,L-Lactamide	1.07 (1.02-1.11)	1.05 (0.97-1.11)	0.030	0.821	
D-Glucosamine	4.77 (1.88-6.01)	5.20 (4.01-5.63)	0.553	<0.001	+
D-Galactosamine	0.95 (0.89-1.00)	0.95 (0.82-1.00)	-0.118	0.369	
D-Mannosamine	1.14 (1.07-1.28)	1.15 (1.03-1.31)	0.154	0.241	
N-Acetyl-D-Glucosamine	5.10 (2.10-6.45)	5.24 (4.10-6.38)	0.438	<0.001	

	Control isolates	Case isolates	Spearman		
	Median (IQR)	Median (IQR)	*rho*	p-value	Sig
N-Acetyl-D-Galactosamine	1.02 (0.96-1.07)	0.99 (0.89-1.05)	-0.119	0.365	
N-Acetyl-D-Mannosamine	2.22 (1.12-3.24)	2.72 (2.07-3.65)	0.559	<0.001	+
Adenine	1.15 (0.96-1.57)	1.35 (1.00-1.70)	0.307	<0.05	
Adenosine	6.26 (5.31-7.17)	6.09 (5.33-7.43)	0.449	<0.001	
Cytidine	6.22 (3.34-7.47)	6.51 (5.63-7.32)	0.539	<0.001	+
Cytosine	1.04 (1.01-1.10)	1.04 (0.99-1.08)	0.101	0.441	
Guanine	2.64 (2.02-3.58)	2.33 (1.93-3.26)	0.002	0.987	
Guanosine	1.27 (1.16-1.78)	1.29 (1.07-1.87)	-0.146	0.267	
Thymine	1.03 (0.99-1.11)	1.00 (0.92-1.12)	-0.040	0.762	
Thymidine	1.02 (0.95-1.09)	0.98 (0.88-1.11)	0.102	0.439	
Uracil	1.00 (0.96-1.02)	0.98 (0.88-1.04)	-0.141	0.282	
Uridine	1.04 (0.98-1.11)	1.03 (0.89-1.10)	0.006	0.963	
Inosine	1.05 (1.00-1.10)	1.03 (0.88-1.12)	-0.110	0.405	
Xanthine	0.98 (0.90-1.13)	0.98 (0.87-1.06)	-0.028	0.833	
Xanthosine	1.14 (1.02-1.42)	1.14 (1.03-1.41)	0.459	<0.001	
Uric Acid	1.48 (1.23-1.78)	1.40 (1.26-1.58)	0.077	0.561	
Alloxan	1.17 (1.07-1.25)	1.17 (1.01-1.20)	0.223	0.086	
Allantoin	1.01 (0.97-1.05)	1.01 (0.93-1.04)	0.208	0.111	
Parabanic Acid	1.03 (0.97-1.06)	1.03 (0.95-1.06)	0.223	0.087	
D,L-α-Amino-N-Butyric Acid	0.86 (0.78-0.93)	0.81 (0.73-0.87)	-0.253	0.051	
γ-Amino-N-Butyric Acid	1.04 (1.01-1.12)	1.04 (0.99-1.21)	0.436	<0.001	
ε-Amino-N-Caproic Acid	1.00 (0.98-1.04)	1.03 (0.96-1.07)	0.391	<0.01	
D,L-a-Amino-Caprylic Acid	1.44 (1.26-1.88)	1.43 (1.15-1.87)	-0.061	0.642	
δ-Amino-N-Valeric Acid	1.07 (0.99-1.32)	1.08 (1.00-1.26)	0.556	<0.001	+
α-Amino-N-Valeric Acid	0.86 (0.81-0.92)	0.84 (0.76-0.90)	-0.294	<0.05	
Ala-Asp	4.40 (1.30-5.41)	4.69 (3.29-5.63)	0.629	<0.001	+
Ala-Gln	4.93 (1.25-6.29)	5.15 (4.28-6.30)	0.578	<0.001	+
Ala-Glu	4.38 (1.37-5.56)	4.77 (3.81-5.93)	0.653	<0.001	+
Ala-Gly	3.45 (1.11-4.42)	4.02 (2.67-4.72)	0.672	<0.001	+
Ala-His	2.77 (1.03-3.61)	3.17 (2.14-4.06)	0.668	<0.001	+
Ala-Leu	2.88 (1.08-3.72)	3.07 (1.83-3.95)	0.651	<0.001	+
Ala-Thr	3.04 (1.07-3.60)	3.19 (2.25-4.13)	0.654	<0.001	+
Gly-Asn	4.16 (1.45-5.27)	4.61 (3.38-5.68)	0.629	<0.001	+
Gly-Gln	4.50 (1.43-5.77)	4.70 (3.42-5.94)	0.594	<0.001	+
Gly-Glu	1.57 (1.16-3.13)	2.28 (1.29-2.90)	0.493	<0.001	+
Gly-Met	1.64 (1.01-2.73)	2.35 (1.22-3.21)	0.650	<0.001	+
Met-Ala	2.07 (1.03-3.25)	2.50 (1.35-3.58)	0.645	<0.001	+
PM4 – Phosphorus sources					
Phosphate	1.13 (1.01-2.40)	2.10 (1.14-3.26)	0.412	<0.01	
Pyrophosphate	3.35 (2.35-4.03)	3.26 (2.74-4.20)	0.070	0.598	
Trimeta-phosphate	1.05 (1.00-2.13)	1.59 (1.15-2.89)	0.406	<0.01	
Tripoly-phosphate	1.04 (0.98-1.85)	1.40 (1.04-2.63)	0.417	<0.01	
Triethyl Phosphate	1.00 (0.99-1.04)	1.03 (1.01-1.08)	0.137	0.296	

	Control isolates	Case isolates	Spearman		
	Median (IQR)	Median (IQR)	*rho*	p-value	Sig
Hypophosphite	1.00 (0.97-1.02)	0.99 (0.96-1.01)	-0.275	<0.05	
Adenosine-2'-monophosphate	3.43 (1.14-4.80)	4.70 (3.85-5.41)	0.347	<0.05	
Adenosine-3'-monophosphate	5.38 (1.93-6.79)	6.30 (4.99-7.00)	0.484	<0.001	+
Adenosine-5'-monophosphate	5.01 (2.50-6.53)	5.25 (4.53-6.58)	0.390	<0.01	
Adenosine-2',3'-cyclic monophosphate	4.77 (1.33-6.02)	5.77 (4.48-6.34)	0.456	<0.001	
Adenosine-3',5'-cyclic monophosphate	1.17 (1.14-1.21)	1.20 (1.17-1.24)	0.253	0.051	
Thio-phosphate	1.37 (1.00-2.71)	2.15 (1.04-3.16)	0.552	<0.001	+
Dithio-phosphate	1.22 (1.06-2.36)	1.85 (1.08-2.87)	0.581	<0.001	+
D,L-α-Glycerol Phosphate	2.77 (1.12-4.04)	3.63 (3.15-4.80)	0.613	<0.001	+
ß-Glycerol Phosphate	1.77 (1.06-3.24)	2.95 (1.70-4.05)	0.587	<0.001	+
Carbamyl Phosphate	1.17 (1.07-2.34)	2.03 (1.20-2.98)	0.379	<0.01	
D-2-Phospho-Glyceric Acid	2.37 (1.25-3.32)	2.90 (1.74-3.73)	0.602	<0.001	+
D-3-Phospho-Glyceric Acid	3.04 (1.85-4.41)	3.99 (3.03-4.67)	0.489	<0.001	+
Guanosine-2'-monophosphate	2.46 (1.10-3.96)	3.72 (2.65-4.56)	0.455	<0.001	
Guanosine-3'-monophosphate	5.06 (1.55-6.75)	6.13 (4.87-7.15)	0.585	<0.001	+
Guanosine-5'-monophosphate	2.85 (1.12-5.61)	3.91 (2.73-5.63)	0.491	<0.001	+
Guanosine-2',3'-cyclic monophosphate	2.82 (1.24-4.75)	4.52 (2.58-5.35)	0.483	<0.001	+
Guanosine-3',5'-cyclic monophosphate	1.26 (1.20-1.30)	1.25 (1.22-1.29)	-0.055	0.675	
Phosphoenol Pyruvate	2.76 (1.40-4.27)	3.71 (2.37-4.50)	0.493	<0.001	+
Phospho-Glycolic Acid	1.20 (1.05-2.73)	2.38 (1.45-3.31)	0.411	<0.01	
D-Glucose-1-Phosphate	4.70 (1.32-5.60)	5.32 (4.30-5.98)	0.531	<0.001	+
D-Glucose-6-Phosphate	4.59 (1.89-5.62)	5.64 (4.13-6.04)	0.619	<0.001	+
2-Deoxy-D-Glucose 6-Phosphate	1.11 (1.08-1.14)	1.13 (1.11-1.20)	0.066	0.616	
D-Glucos-amine-6-Phosphate	3.61 (2.46-4.97)	4.59 (3.79-5.47)	0.630	<0.001	+
6-Phospho-Gluconic Acid	1.23 (1.11-1.50)	1.32 (1.26-1.77)	0.294	<0.05	
Cytidine-2'-monophosphate	5.43 (2.38-7.11)	6.22 (4.81-7.12)	0.591	<0.001	+
Cytidine-3'-monophosphate	1.14 (1.09-1.20)	1.17 (1.14-1.26)	0.319	<0.05	
Cytidine-5'-monophosphate	3.10 (1.14-4.65)	3.01 (1.72-4.45)	0.362	<0.01	
Cytidine-2',3'-cyclic monophosphate	4.33 (1.29-6.37)	5.51 (4.17-6.40)	0.596	<0.001	+
Cytidine-3',5'-cyclic monophosphate	1.23 (1.20-1.26)	1.26 (1.24-1.31)	0.338	<0.05	
D-Mannose-1-Phosphate	2.93 (1.08-5.15)	4.61 (2.35-5.57)	0.614	<0.001	+
D-Mannose-6-Phosphate	3.69 (1.58-4.82)	4.74 (3.81-5.37)	0.606	<0.001	+
Cysteamine-S-Phosphate	1.20 (1.06-2.71)	2.27 (1.34-3.52)	0.545	<0.001	+
Phospho-L-Arginine	1.62 (1.09-3.45)	2.98 (1.81-3.94)	0.532	<0.001	+
O-Phospho-D-Serine	1.15 (1.12-1.58)	1.43 (1.13-1.99)	0.440	<0.001	
O-Phospho-L-Serine	1.27 (1.16-1.72)	1.33 (1.24-2.24)	0.346	<0.05	
O-Phospho-L-Threonine	1.45 (1.17-1.79)	1.72 (1.35-2.10)	0.481	<0.001	+
Uridine-2'-monophosphate	2.87 (1.31-4.55)	4.02 (3.28-5.17)	0.500	<0.001	+
Uridine-3'-monophosphate	5.10 (1.92-6.63)	5.94 (4.56-6.52)	0.550	<0.001	+

	Control isolates	Case isolates	Spearman		
	Median (IQR)	Median (IQR)	*rho*	p-value	Sig
Uridine-5'-monophosphate	2.56 (1.12-4.68)	3.61 (2.67-4.87)	0.519	<0.001	+
Uridine-2',3'-cyclic monophosphate	5.20 (1.57-6.84)	6.11 (4.76-6.92)	0.574	<0.001	+
Uridine-3',5'-cyclic monophosphate	1.13 (1.09-1.15)	1.14 (1.12-1.18)	0.279	<0.05	
O-Phospho-D-Tyrosine	2.16 (1.07-3.50)	2.98 (1.91-3.90)	0.425	<0.01	
O-Phospho-L-Tyrosine	1.62 (1.06-3.37)	2.93 (2.12-3.69)	0.442	<0.001	
Phosphocreatine	1.21 (1.08-2.26)	2.07 (1.24-3.28)	0.399	<0.01	
Phosphoryl Choline	1.13 (1.08-1.22)	1.18 (1.14-1.33)	0.266	<0.05	
O-Phosphoryl-Ethanolamine	1.23 (1.14-1.31)	1.27 (1.22-1.46)	0.385	<0.01	
Phosphono Acetic Acid	1.24 (1.17-1.29)	1.23 (1.19-1.29)	0.009	0.946	
2-Aminoethyl Phosphonic Acid	1.43 (1.25-2.28)	2.25 (1.40-2.84)	0.416	<0.01	
Methylene Diphosphonic Acid	1.70 (1.65-1.92)	1.87 (1.67-2.02)	0.013	0.920	
Thymidine-3'-monophosphate	2.03 (1.19-3.70)	3.74 (2.83-4.36)	0.487	<0.001	+
Thymidine-5'-monophosphate	2.84 (1.33-5.03)	3.92 (3.35-5.03)	0.537	<0.001	+
Inositol Hexaphosphate	2.35 (1.25-3.63)	2.74 (1.25-3.32)	0.042	0.752	
Thymidine 3',5'-cyclic monophosphate	1.20 (1.17-1.25)	1.21 (1.19-1.25)	0.130	0.323	
PM4 – Sulfur sources					
Sulfate	1.05 (1.01-1.12)	1.05 (1.02-1.27)	0.293	<0.05	
Thiosulfate	1.11 (1.01-1.53)	1.36 (1.11-1.52)	0.459	<0.001	
Tetrathionate	1.07 (1.02-1.25)	1.20 (1.05-1.50)	0.360	<0.01	
Thiophosphate	1.21 (1.02-1.79)	1.38 (1.15-1.76)	0.547	<0.001	+
Dithiophosphate	1.16 (1.01-1.70)	1.44 (1.08-1.70)	0.506	<0.001	+
L-Cysteine	1.11 (1.04-1.38)	1.26 (1.05-1.40)	0.343	<0.05	
D-Cysteine	1.33 (1.04-1.81)	1.59 (1.25-1.74)	0.597	<0.001	+
L-Cysteinyl-Glycine	1.17 (1.05-1.61)	1.41 (1.26-1.59)	0.590	<0.001	+
L-Cysteic Acid	1.08 (1.02-1.28)	1.10 (1.02-1.26)	0.389	<0.01	
Cysteamine	1.07 (1.01-1.11)	1.05 (1.03-1.11)	0.302	<0.05	
L-Cysteine Sulfinic Acid	1.53 (1.09-2.03)	1.60 (1.25-1.82)	0.571	<0.001	+
N-Acetyl-L-Cysteine	1.01 (0.91-1.04)	0.98 (0.93-1.01)	-0.496	<0.001	+
S-Methyl-L-Cysteine	1.01 (0.95-1.04)	1.00 (0.93-1.06)	-0.276	<0.05	
Cystathionine	1.22 (1.07-1.48)	1.42 (1.18-1.59)	0.622	<0.001	+
Lanthionine	1.50 (1.08-1.85)	1.60 (1.36-1.85)	0.588	<0.001	+
Glutathione	1.32 (1.11-1.82)	1.60 (1.25-1.78)	0.552	<0.001	+
D,L-Ethionine	0.97 (0.62-1.08)	0.82 (0.57-1.03)	-0.511	<0.001	+
L-Methionine	1.34 (1.12-1.57)	1.51 (1.32-1.66)	0.496	<0.001	+
D-Methionine	1.41 (1.11-1.81)	1.69 (1.34-1.96)	0.621	<0.001	+
Glycyl-L-Methionine	1.44 (1.14-1.85)	1.60 (1.46-2.00)	0.591	<0.001	+
N-Acetyl-D,L-Methionine	1.29 (1.11-1.61)	1.54 (1.35-1.69)	0.568	<0.001	+
L-Methionine Sulfoxide	1.33 (1.07-1.61)	1.45 (1.28-1.71)	0.572	<0.001	+
L-Methionine Sulfone	0.99 (0.63-1.12)	0.80 (0.50-1.02)	-0.464	<0.001	
L-Djenkolic Acid	1.40 (1.07-1.60)	1.45 (1.16-1.66)	0.557	<0.001	+
Thiourea	1.08 (1.05-1.23)	1.14 (1.08-1.27)	0.508	<0.001	+
1-Thio-ß-D-Glucose	1.08 (1.04-1.12)	1.07 (1.04-1.12)	-0.302	<0.05	

	Control isolates	Case isolates	Spearman		
	Median (IQR)	Median (IQR)	*rho*	p-value	Sig
D,L-Lipoamide	1.08 (1.04-1.12)	1.09 (1.03-1.12)	0.151	0.249	
Taurocholic Acid	1.09 (1.04-1.15)	1.10 (1.05-1.16)	-0.130	0.324	
Taurine	1.04 (0.99-1.09)	1.01 (0.94-1.07)	-0.223	0.087	
Hypotaurine	1.07 (1.04-1.10)	1.03 (0.97-1.09)	-0.107	0.417	
p-Amino Benzene Sulfonic Acid	1.09 (1.04-1.13)	1.08 (0.99-1.12)	-0.298	<0.05	
Butane Sulfonic Acid	1.06 (0.97-1.15)	1.04 (0.93-1.10)	-0.295	<0.05	
2-Hydroxyethane Sulfonic Acid	1.07 (1.01-1.11)	1.03 (0.97-1.07)	-0.228	0.080	
Methane Sulfonic Acid	1.07 (1.01-1.10)	1.02 (0.97-1.07)	-0.212	0.105	
Tetra-methylene Sulfone	1.07 (0.95-1.14)	1.03 (0.96-1.09)	-0.344	<0.05	

Sig: significance after Bonferroni correction (empty cells: not significant, +: significant).

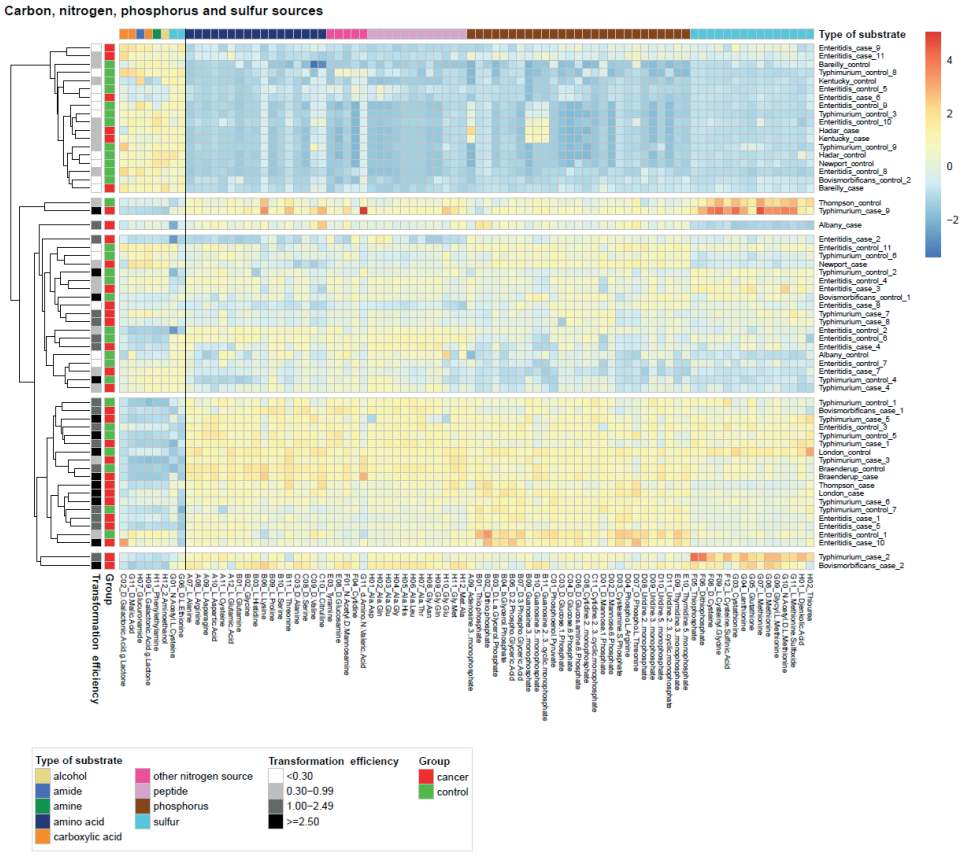


Figure S5. Heatmap of scaled utilization scores of the 60 nontyphoidal *Salmonella* (NTS) isolates for the sources with a significant positive or inverse correlation with transformation efficiency. Left eight columns depict the sources with inverse correlation (left of black line), right 76 columns depict the source with a positive correlation. NTS isolates are clustered based their utilization scores using the average linkage method.

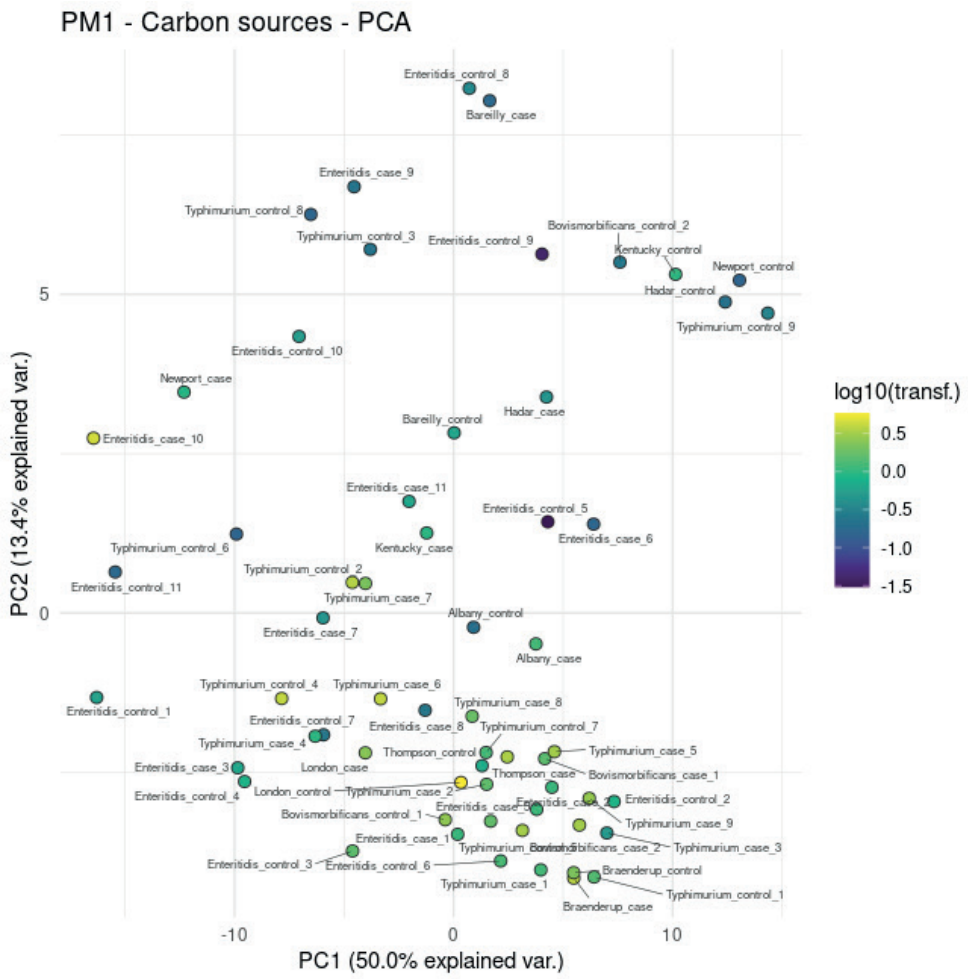


Figure S6. Principal component analysis (PCA) plot of the carbon source utilization of 60 nontyphoidal *Salmonella* isolates. The color scale of the dots reflects the log transformation efficiency. Principal component 1 (PC1) and PC2 collectively accounted for 63.4% of the variance in the nutrient utilization data. Transf: transformation efficiency.

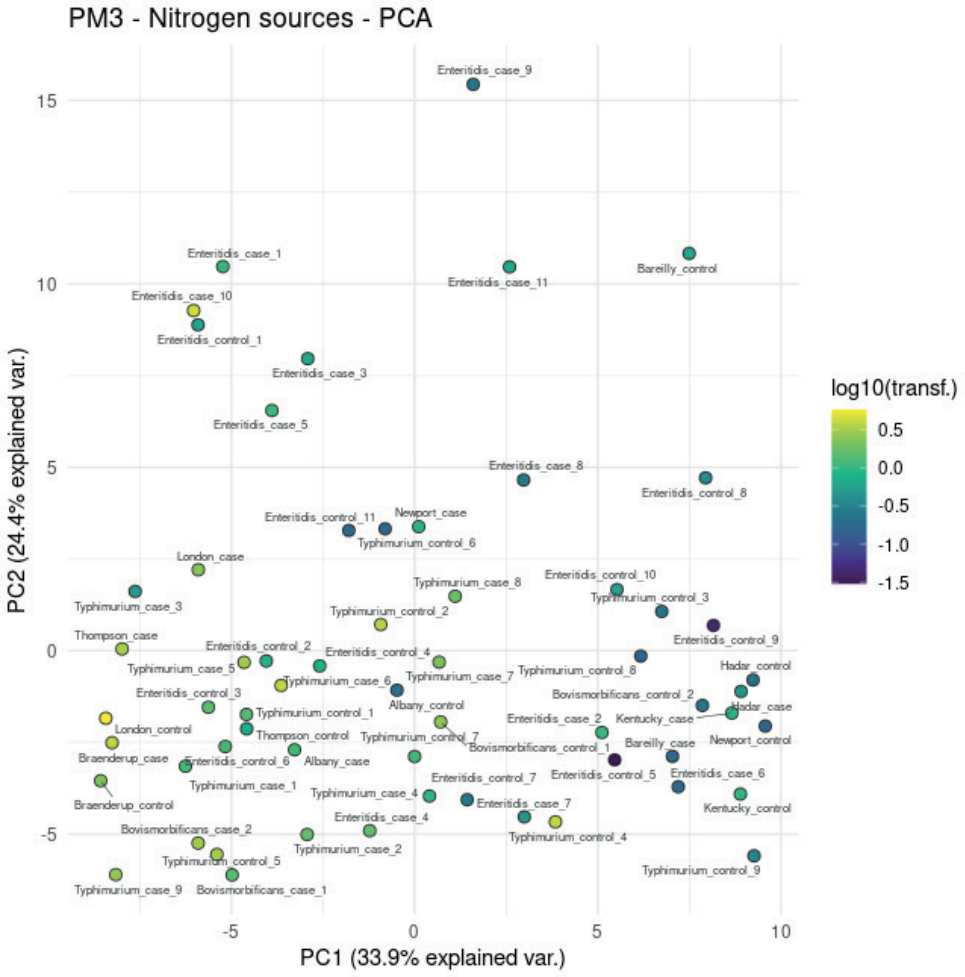


Figure S7. Principal component analysis (PCA) plot of the nitrogen source utilization of 60 nontyphoidal *Salmonella* isolates. The color scale of the dots reflects the log transformation efficiency. Principal component 1 (PC1) and PC2 collectively accounted for 58.3% of the variance in the nutrient utilization data. Transf.: transformation efficiency.

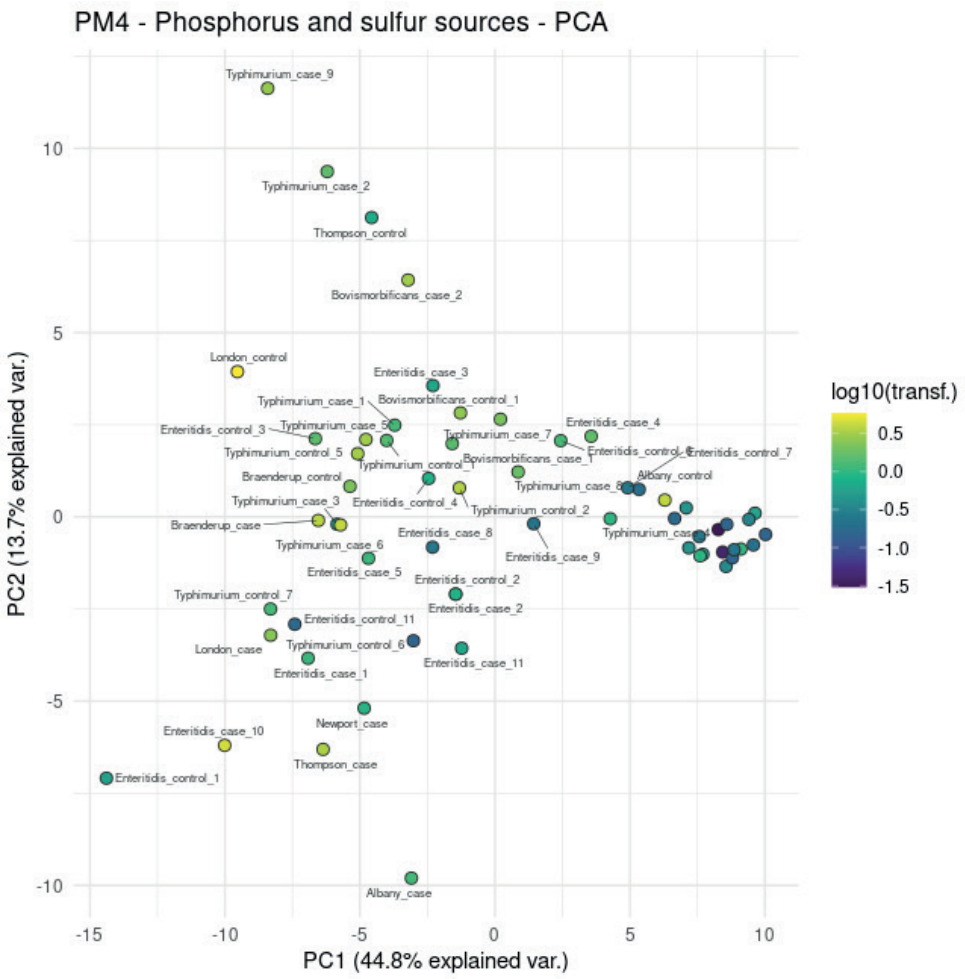


Figure S8. Principal component analysis (PCA) plot of the phosphorus and sulfur source utilization of 60 nontyphoidal *Salmonella* isolates. The color scale of the dots reflects the log transformation efficiency. Principal component 1 (PC1) and PC2 collectively accounted for 58.5% of the variance in the nutrient utilization data. Transf.: transformation efficiency.