UNIVERSITÀ DI PARMA

UNIVERSITÀ DEGLI STUDI DI PARMA

PhD program in Biotechnology and Life Sciences

XXXI Cycle

Study of genetic determinants involved in *Salmonella* Derby host adaptation

Coordinator:
Prof. Simone Ottonello

Tutor:
Gabriele Casadei

PhD student:
Martina Tambassi

Years 2015/2018

Index

Abstract

Salmonella is the second most common cause of food-borne disease in Europe. The majority of Salmonellae causing disease in human and domestic animals belong to *S. enterica* subspecies *enterica*, which is divided in approximately 1,500 serovars, based on their unique somatic and flagellar antigenic formulas. Serovars can exhibit different host-tropism: generalist serovars are able to infect a variety of unrelated hosts, host-adapted serovars are mainly isolated from a specific host, and specialists serovars are restricted to a single host. Epidemiological evidence supports various levels of host adaptation even among isolates belonging to the same serovar. A number of genetic features have been consistently associated to serovars with a specific host range, but the exact mechanism of host adaptation still remains elusive. Discovering the genetic determinants involved in the ability of *S. enterica* serovars to infect specific hosts is critically important to assess the risk of pathogen transmission along the food-chain. Data from the Emilia-Romagna Reference Centre for Enteropathogens Surveillance show that *S. enterica* serovar Derby is mainly prevalent in swine (28.2%), while it is only rarely found in human (2.6%). *S*. Derby isolates were genotyped by Pulsed-Field Gel Electrophoresis (PFGE). Different proportion of human isolates were observed for the two most prevalent PFGE profiles in swine: SXB_BS.0204 is significantly less frequently isolated in humans (0.9%) than in swine (9.4%), while SXB_BS.0056 is isolated in humans (13%) and in swine (9.7%) at a similar frequency. Based on these data, the aim of my thesis was to identify the genetic determinants responsible for different distribution in humans of the two PFGE profiles.

Specifically, we evaluated whether the different prevalence in humans of SXB_BS.0204 and SXB_BS.0056 isolates were due to differences in virulence. Infection assays in the *Caenorhabditis elegans* model did not detect differences between isolates belonging to distinct PFGE profiles in their ability to kill the nematode. We also performed invasion and replication assays in human INT-407 and swine IPEC-J2 epithelial cell lines. We found that invasion and replication efficiencies in human cells of SXB_BS.0056 isolates were up to 4 logs higher than those of SXB_BS.0204 isolates. In swine cells, SXB_BS.0056 isolates were more virulent than SXB_BS.0204 isolates, but differences were significantly reduced to just one log. Isolates were thus defined virulent (V) and non-virulent (NV) based on their infection efficiency. Whole genomes of 45 isolates were sequenced and core single nucleotide polymorphisms (SNPs) were used as input for generating a phylogenetic tree. V and NV isolates cluster separately, suggesting different evolutionary paths. We assessed if

exclusive genes of V isolates were responsible for their virulence phenotype. Mutants deleted for plasmids and genomic regions found only in V isolates showed no changes in infection efficiency compared to the respective wild types. To evaluate the role of different gene allelic variants in virulence, we searched for non-synonymous SNPs that discriminated V from NV isolates. Two allelic variants of *hilD* were detected: HilD is the activator of *Salmonella* Pathogenic Island (SPI) 1, which encodes for the Type 3 Secretion System (T3SS) responsible for invasion of epithelial cells. In human cells, a V isolate (ER1175 WT) carrying the *hilD* NV allele (*hilD_nv*) has the same low infection efficiency of NV isolates. Accordingly, a NV isolate (ER278 WT) carrying the V *hilD* allele (ER278::*hilD_v*) shows the same high infection efficiency of V isolates. Both strains deleted individually for *hilD* and *invA* (encoding for an essential component of T3SS) infect human cells at the same level of NV isolates. In swine cells, the introduction of *hilD_nv* as well as the *hilD* deletion in ER1175 WT did not cause changes in the virulence phenotype, whereas the *invA* deletion triggered a decrease in virulence. No differences were found among ER278 WT, ER278Δ*hilD* and ER278Δ*invA* in swine cells, but *hilD_v* insertion causes an increase in both invasion and replication compared to ER278 WT. The two *hilD* allelic variants thus explain differences in virulence between V and NV isolates in human cells, but not in swine cells. To assess differences in SPI-1 gene expression caused by *hilD* allelic variants, an RNA-seq-based transcriptomic analysis was performed by extracting RNA from strains cultured in SPI-1-inducing conditions *in vitro*. There were no differences in gene expression between strains carrying *hilD_nv* and the respective *hilD*-deleted strains, whereas SPI-1 genes were upregulated in both ER1175 and ER278 carrying *hilD_v*. These results demonstrate that *hilD_nv* produces a non-functional protein. The same results were obtained analyzing the expression of SPI-1 genes in ER1175 WT, ER1175::*hilD_nv* and ER1175Δ*hilD* during infection of human cells. These three strains and ER278 WT were also tested in the streptomycin-treated mouse model of colitis of *Salmonella* infection to assess if *hilD* variants cause differences in *in vivo* host-bacterial interaction. No differences among these strains were found in cecal bacterial loads as well as in the expression of host genes correlated with inflammation (*Nos2* and *Cxcl1*), which was not significantly different from uninfected mice. Thus, these strains do not seem to induce gut inflammation in mice. No differences in systemic dissemination were observed between ER278 WT, ER1175 WT, and ER1175::*hilD_nv*. In contrast, ER1175Δ*hilD* replicated in liver and spleen at higher level than ER1175 WT, suggesting a negative effect of HilD on *S*. Derby dissemination. Together, this study has provided new information about mechanisms of host adaptation adopted by *S.* Derby.

Riassunto

Salmonella è la seconda causa più comune di zoonosi in Europa. La maggior parte delle Salmonelle che causano malattia nell'uomo e negli animali domestici appartengono a *S. enterica sottospecie enterica*, divisa in circa 1500 sierotipi in base ai loro specifici antigeni somatici e flagellari. Sierotipi diversi possono esibire un diverso spettro d'ospite: sono stati identificati sierotipi generalisti, in grado di infettare più specie, sierotipi ospite adattati, maggiormente associati a un ospite, e sierotipi specialisti, ristretti a un unico ospite. Evidenze epidemiologiche hanno inoltre evidenziato diversi livelli di adattamento all'ospite anche tra isolati appartenenti allo stesso sierotipo. Anche se alcune caratteristiche genetiche sono state associate a sierotipi con uno spettro di specie infettate ben determinato, i meccanismi coinvolti nell'adattamento all'ospite di questi patogeni rimangono ancora largamente incompresi. Individuare i determinanti genici coinvolti nella capacità di infettare specifici ospiti è determinante nel valutare il rischio di trasmissione del patogeno lungo la catena alimentare. Dati provenienti dal Centro di Referenza dell'Emilia Romagna per la sorveglianza degli enteropatogeni hanno evidenziato che il sierotipo Derby è principalmente isolato in suino (28%) mentre è solo raramente trovato in uomo (2.6%). Gli isolati di *S*. Derby sono stati genotipizzati mediante Pulsed-Field Gel Electrophoresis (PFGE) ed è emerso che i due profili PFGE più prevalenti in suino hanno diversa distribuzione in uomo: SXB_BS.0204 è significativamente meno isolato in uomo (0.9%) che in suino (9.4%), mentre SXB_BS.0056 è proporzionalmente isolato allo stesso modo in uomo (13%) e in suino (9.7%). Lo scopo della tesi è stato identificare i determinanti genici responsabili della diversa distribuzione in uomo dei due profili PFGE. È stato valutato se le differenze osservate fossero dovute a diversità tra gli isolati SXB_BS.0204 e SXB_BS.0056 nel fenotipo di virulenza. Saggi di infezione su *C. elegans* non hanno evidenziato alcuna differenza tra isolati con pulsotipi diversi nella capacità di uccidere il nematode. Sono stati quindi eseguiti saggi di invasione e replicazione in linee cellulari epiteliali umane (INT-407) e suine (IPEC-J2). Nelle cellule umane l'efficienza di invasione e replicazione degli isolati SXB_BS.0056 è fino a 4 logaritmi più alta di quella degli isolati SXB_BS.0204. Nelle cellule suine gli isolati SXB_BS.0056 sono più virulenti degli isolati SXB_BS.0204, ma la differenza è ridotta a un logaritmo. Gli isolati sono stati definiti virulenti (V) e non virulenti (NV) in base alla loro efficienza di infezione. 45 isolati sono stati sequenziati per generare un albero filogenetico basato sui polimorfismi a singolo nucleotide (SNPs). Gli isolati NV clusterizzano separatamente dai V, suggerendo diverse strade evolutive. È stato verificato se i geni esclusivi degli isolati V fossero

responsabili del loro maggiore capacità di infettare. I mutanti deleti per i plasmidi e le regioni genomiche individuate solo negli isolati V non hanno mostrato fenotipo di virulenza diverso dai rispettivi isolati wild type. È stato quindi valutato se specifiche varianti alleliche discriminanti gli isolati NV dai V fossero implicate nella virulenza. Sono stati individuati due alleli di HilD, l'attivatore dell'Isola di Patogenicità di *Salmonella* (SPI) 1 che codifica per il sistema di secrezione di tipo 3 (T3SS) responsabile dell'invasione delle cellule epiteliali. In cellule umane un isolato V (ER1175 WT) recante l'allele di *hilD* degli isolati NV (*hilD_nv*) ha un fenotipo non virulento. Allo stesso modo un isolato NV (ER278 WT) con l'allele di *hilD* degli isolati V (*hilD_v*) ha fenotipo virulento. Entrambi i ceppi deleti per *hilD* infettano le cellule umane allo stesso livello degli isolati NV. La delezione di *hilD* così come la sua sostituzione con *hilD_nv* in ER1175 WT non causano cambiamenti nell'efficienza di infezione delle cellule suine, mentre la delezione di *invA* (codificante un componente essenziale del T3SS) diminuisce il livello di virulenza. Non sono state riscontrate differenze tra ER278 WT, ER278Δ*hilD* e ER278Δ*invA* in cellule suine, ma l'inserzione di *hilD_v* causa un aumento di invasione e replicazione del mutante rispetto a ER278 WT. Le due varianti alleliche di *hilD* sono quindi responsabili delle differenze osservate nelle cellule umane, ma non spiegano quelle osservate nelle cellule suine. Per valutare se la presenza di varianti alleliche diverse di *hilD* causi una diversa espressione della SPI-1, è stata eseguita un'analisi trascrittomica sull'RNA estratto in condizioni *in vitro* determinanti l'upregolazione di SPI-1. Non sono state riscontrate differenze nell'espressione genica tra i ceppi recanti *hilD_nv* e i rispettivi mutanti deleti in *hilD*, mentre i geni della SPI-1 sono upregolati sia in ER278 che ER1175 recanti *hilD_v*. I risultati di transcrittomica *in vitro* provano che *hilD_nv* esprime una proteina non funzionale. Gli stessi risultati sono stati ottenuti analizzando l'espressione dei geni della SPI-1 in ER1175 WT, ER1175::*hilD_nv* e ER1175Δ*hilD* durante infezione delle cellule umane. La virulenza di questi 3 ceppi e di ER278 WT è stata testata sul modello murino di infezione di *Salmonella*. I ceppi testati si accumulano nell'intestino ceco a livelli simili. Non sono state riscontrate differenze tra topi infettati nel livello di espressione di geni correlati con l'infiammazione (*Nos2* and *Cxcl1*), espressi allo stesso livello dei topi di controllo non infettati. *S*. Derby sembra quindi non causare infiammazione intestinale. ER278 WT, ER1175 WT e ER1175::*hilD_nv* mostrano livelli di disseminazione non significativamente diversi. ER1175Δ*hilD* si accumula in fegato e milza a un livello significativamente maggiore rispettivo al ceppo wild type, suggerendo un effetto negativo di *hilD* sulla disseminazione di *S*. Derby. Questo studio ha fornito nel suo insieme nuove informazioni riguardo il meccanismo di adattamento all'ospite di *S*. Derby.

1. Introduction

1.1. The Genus *Salmonella*

1.1.1. Taxonomy

The Genus *Salmonella* was discovered in 1885 by Theobald Smith, the assistant of the veterinary pathologist Daniel Elmer Salmon. Searching for the cause of common hog cholera, Smith isolated what was called *Salmonella cholera-suis*, because his head Salmon claimed credit for the discovery (1). The genus *Salmonella* belongs to the Enterobacteriaceae family and it consists of Gram-negative, non-spore forming, rod-shaped, facultative anaerobes bacteria which show peritrichous flagella motility. This genus is composed by two species, *S. enterica* and *S. bongori*. *S. enterica* is divided into six subspecies: *S. enterica* subsp. *enterica* (I), *S. enterica* subsp. *salamae* (II), *S. enterica* subsp. *arizonae* (IIIa), *S. enterica* subsp. *diarizonae* (IIIb), *S. enterica* subsp. *houtenae* (IV), and *S. enterica* subsp. *indica* (VI). The majority of *Salmonella*e that cause disease in human and domestic animals belong to *S. enterica* subsp. *enterica* (I). All the other subspecies and *S. bongori* are more typically isolated from cold-blooded animals (2).

Salmonella strains can be classified based on the somatic (O) antigens into 67 serogroups and into 2,579 serotypes or serovars when strains are classified even by flagellar (H) antigens. This classification is based on the Kauffmann-White-Le Minor scheme, continuously updated by the World Health Organization (WHO) Collaborating Centre for Reference and Research on *Salmonella* at the Pasteur Institute, Paris, France (2). 1,531 serovars are recognized to belong to subsp. *enterica*. The subspecies name (subsp. *enterica*) does not need to be indicated as only serovars of this subspecies bear a name. Serovars of other subspecies of *S. enterica* and those of *S. bongori* are designated only by their antigenic formula. The complete nomenclature, for example *S. enterica* subsp. *enterica* serovar Typhimurium, can be shorten to *S.* Typhimurium.

Salmonella enterica subsp. *enterica* is commonly divided also in typhoidal and non-typhoidal

Salmonella (NTS) strains. *Salmonella* Typhi and Paratyphi (the typhoidal serovars) are human-restricted pathogens that cause enteric (typhoid) fever, a systemic disease characterized by fever, headache, abdominal pain, transient diarrhea or constipation. The NTS strains are pathogens that can infect a broad range of hosts, including humans and several animals. They usually cause acute self-limiting gastroenteritis localized to the terminal ileum and colon, but they can also cause bacteremia and systemic infection in immunocompromised hosts, in very young and older individuals and occasionally in healthy adult humans and animals (3).

1.1.2. Typing methods

It is increasingly important for the epidemiology surveillance systems to be able to distinguish *Salmonella* isolates, in order to detect outbreaks and their relative source of infection. *Salmonella* isolates can be differentiated from one another by a wide variety of techniques other than serotyping, including phage typing and several molecular methods (reviewed in 4).

Phage typing utilizes the selective ability of bacteriophages to infect certain strains of *Salmonella* due to differences in phage receptors present on the surface of the bacterium. A phage type designation is assigned based on the array of typing phages that are able to lyse the cells and form plaques in the bacterial lawns.

Molecular typing techniques include Pulsed-field gel electrophoresis (PFGE). This technique is based on the separation of genomic DNA, previously digested with unique restriction enzymes, on a gel matrix under an electric field that periodically changes direction. This electric field permits analysis of bacterial DNA fragments over an order of magnitude larger than those analyzable by standard electrophoresis. The banding patterns generated by PFGE are analyzed to assign to each isolate a PFGE profile (or pulsotype). Cluster analysis of the pulsotypes are then carried out based on the unweighted pair group method with arithmetic averages (UPGMA). PFGE is considered worldwide as the “gold standard” fingerprinting method for *Salmonella* typing.

Other molecular techniques developed for isolate typing are based on amplification of specific and nonspecific target. Like PFGE, these techniques generate characteristic banding patterns which are compared to each other to evaluate the degree of similarity (4). These methods include:

- Amplified fragment length polymorphisms (AFLP): extracted DNA is digested with one or more enzymes, then linkers are ligated to the restriction fragments and used as primers binding site to allow fragments amplification;
- Random amplified polymorphic DNA PCR (RAPDPCR) and arbitrary primed PCR (AP-PCR): these techniques use random primers which amplify variably sized amplicons;
- Repetitive element PCR (Rep-PCR): PCR primers for this technique are designed on repetitive elements. When two repeated elements are located in close proximity to one another, the intervening region, that can be differentially sized, is amplified.
- Multiple Locus Variable-Number Tandem Repeat Analysis (MLVA): this method detects by PCR the copy numbers of repeated DNA sequences that are dispersed throughout the bacterial genome. As the number of these tandem repeats in a particular locus varies from strain to strain, they are known as variable-number tandem repeats or VNTRs. For a particular bacterium, VNTRs may be present in multiple loci or regions.

AFLP, RAPDPCR and Rep-PCR are no longer used as routine typing methods, replaced by MLVA and other molecular methods based on sequencing technology.

Multilocus sequence typing (MLST) is a molecular typing method based on sequencing of specific housekeeping genes and comparison of DNA sequence polymorphisms to differentiate strains. The different sequences present within a bacterial species of each housekeeping gene are considered distinct alleles and are associated to a specific integer. Each isolate of a species is therefore unambiguously characterized by an allelic profile or sequence type (ST), which corresponds to a series of integers associated to the alleles of the housekeeping loci.

Nowadays, several surveillance reference laboratories all around the world are working to transition from the classical typing methods to Whole Genome Sequencing (WGS), which allows to perform the so called “genomic epidemiology” (reviewed in 5 and 6). Because it provides the whole genetic information of an organism, WGS is considered the most advanced solution for epidemiologic investigation of foodborne pathogens. The high-throughput sequencing (called Next Generation Sequencing, NGS) technology more commonly used in the field of the genomic epidemiology is Illumina, followed by Ion Torrent and Pacific Biosciences (PacBio) sequencing. Illumina and Ion Torrent produce short reads, up to 300 bp and 400 bp, respectively. The Illumina platform is based on a sequencing-by-synthesis technology, using reversible (fluorescent) terminators, whereas the Ion Torrent technology relies on so-called semiconductor sequencing. PacBio technology performs single-molecule real-time (SMRT) sequencing, producing long reads, up to 60 kb.

The most diffused bioinformatics approach to analyze the genomic data obtained by NGS for epidemiology is the reference sequence–based method. Reads from each analyzed isolate are mapped to a reference genome to identify genomic variants, which include single nucleotide polymorphisms (SNPs) and small insertions and deletions (indels). Most of the available microbial pipelines for surveillance and outbreak investigation, such as Snippy (7) and CFSAN SNP Pipeline (8), focus on SNPs. In the reference sequence–based approach, only genomic regions that are present in the reference sequence and in all genomes analyzed (core genome) can be queried for putative genomic variants. The SNP matrix obtained is used as input for phylogenetic analysis, in order to reconstruct the evolutionary history of a bacterial population. The algorithms mostly used for the construction of phylogenetic trees are maximum-likelihood and Bayesian methods. They rely on different principles: the maximum-likelihood algorithm attempts to find the tree that is most likely to have produced the observed data (given a model of molecular evolution), whereas

the Bayesian method finds the tree that has the highest posterior probability given the observed data. The most used softwares for maximum-likelihood methods are MEGA6 (9), RaXML (10), FastTree (11), and BEAST (12) is commonly used for Bayesian analyses.

Alternative approaches to analyze genetic relatedness among isolates from partial or total core genome are ribosomal (r) MLST and core genome (cg) MLST. In rMLST comparison of 51 genes encoding ribosomal proteins is performed, whereas a genome wide gene-by-gene comparisons is accomplished for cgMLST. Like in the traditional MLST, distinct alleles identifiers are assigned to a curated set of predefined genes, which ensures interlaboratory reproducibility.

Recently even the accessory genome has been used as additional tool to type bacterial genomes. It has been observed that combining the analysis of accessory genes and plasmids with that of core genome provide a super-resolution view into the epidemiology of bacterial populations (6).

1.2. Pathogenesis model and virulence factors

1.2.1. Overview

Almost all studies related to *Salmonella* pathogenesis focus on *S.* Typhimurium, which is used worldwide as model to understand non-typhoidal *Salmonella* interactions with host cells.

S. Typhimurium infection begins with the ingestion of contaminated food or water. To reach the host gut, *Salmonella* protects itself against gastric acidity activating the acid tolerance response (ATR), which provides an inducible pH-homeostatic function to maintain the intracellular pH at values higher than those of the extracellular environment (13).

Once entered the small intestine, in humans *Salmonella* specifically localizes in the terminal ileum, where it establishes the initial contact with the epithelium. The bacterium can follow different pathways (reviewed in 3 and 14):

- It can induce its own uptake by epithelial cells,
- It can traverse the epithelial barrier through Microfold cells (M cells) that overlie intestinal lymphoid tissues known as Peyer's patches, and translocate to the intestinal lymphoid follicles and the draining mesenteric lymph nodes,
- It can be engulfed by phagocytes encountered in the intestinal lumen, in the submucosa or in the mesenteric lymph nodes.

Soon after adhesion, the invasion of non-phagocytic cells is due to profound cytoskeletal rearrangements. These internal modifications disrupt the normal epithelial brush border and induce the subsequent formation of membrane ruffles that engulf adherent bacteria in large vesicles called *Salmonella*-containing vacuoles (SCVs). Host cells invasion elicits induction of proinflammatory cytokines production, which causes infiltration of neutrophils and fluid into the intestinal lumen, resulting in inflammatory diarrhea. Otherwise, migration of infected phagocytes leads to systemic dissemination of the bacteria via the bloodstream to several additional tissues, such as the spleen and liver, where this pathogen preferentially replicates.

The ability of *Salmonella* to infect and cause disease in a host is due to the presence of several virulence factors encoded by genes located within the chromosome or in virulence plasmids. Chromosomal virulence genes can be found in large cassettes composed of a series of genes and operons called *Salmonella* pathogenicity islands (SPIs). SPIs typically accommodate large clusters of genes that contribute to a particular virulence phenotype, which is generally manifested at a specific time during the course of infection.

In the next paragraphs each step of *Salmonella* pathogenesis is explained together with a description of the virulence factors implicated.

1.2.2. Approach and adhesion to non-phagocytic cells

Once *Salmonella* has reached the intestinal lumen, the initial contact with the epithelium is facilitated by flagella-mediated motility and chemotaxis (15). The bacterial flagellum consists of three structural parts, a basal body, a hook, and a filament. The more than 50 flagellar genes are organized into at least 15 operons, and their expression forms a highly organized cascade called the flagellar regulon (16). Flagellar operons are divided into three classes based on their transcriptional hierarchy. The *flhDC* operon is the only one belonging to flagellar operon class 1 and it is required for expression of all the other flagellar operons. The class 1 operon products, FlhD and FlhC, assemble into an FlhD4C2 heterohexamer, which binds to class 2 promoters to facilitate transcription of the class 2 operons. Class 2 contains operons encoding component proteins of the hook-basal body structure and the flagellum-specific type III export apparatus as well as the flagellum-specific sigma factor σ^{28} (FliA), which is essential for class 3 expression. Class 3 contains operons encoding proteins involved in filament assembly and flagellar motility and chemotaxis.

Intimate attachment between bacteria and the eukaryotic cells, an indispensable prerequisite for invasion, is established also through several virulence determinants other than flagellum: different type of fimbriae, non-fimbrial adhesins, autotransporter proteins and SPI-1–encoded type 3 secretion system (T3SS-1) translocon members. Type 1 fimbriae are encoded by *fim* genes: they are arranged in a single cluster composed by the operon *fimAICDHF*, encoding structural subunits, and three regulatory genes, *fimZ*, *fimY*, and *fimW*. The resulting fimbrial structure mediates adhesion by binding the extracellular matrix glycoprotein laminin through its oligomannoside chains (17). Curli fimbriae genes are arranged into two adjacent operons, *csgBAC* and *csgDEFG* (18). These fimbrie primarily participate to adhesion to eukaryotic cells by binding to the extracellular matrix protein fibronectin (19). The *pef* fimbrial genes are located on the pSLT

virulence plasmid of *S.* Typhimurium. Pef fimbriae specifically bind to the trisaccharide Galβ1-4(Fucα1-3)GlcNAc, also known as the Lewis X (Lex) blood group antigen, expressed mainly by crypt epithelial cells. They remain intact once the inflammatory reaction has been initiated, in contrast to the usual target cells of *Salmonella*, so it is possible that the pathogen binds to human crypt epithelium at later stages of infection (20). *std* operon-encoded fimbriae mediate attachment to human colonic epithelial cell lines by binding to terminal α(1-2)fucose receptors expressed in the cecal mucosa (21). SPI-4 encodes for 6 ORFs, designated *siiA* to *siiF*, involved in adhesion. The SiiC, SiiD, and SiiF proteins form a type 1 secretion system (T1SS) which exports SiiE giant non-fimbrial adhesin, which mediates contact-dependent adhesion to epithelial cells (22). Similarly, the large cell surface protein BapA is also secreted through a T1SS (BapBCD) encoded downstream from the *bapA* gene. Despite the attribution of its main role to biofilm formation, the absence of this protein is also related to lower colonization of the intestinal epithelium (23). The *misL* gene, located within SPI-3, encodes an autotransporter protein, namely a protein which functions as secreted substrate as well as transporter. The outer membrane protein MisL has been reported to bind to fibronectin, an extracellular matrix component, and hence to promote colonization of intestinal epithelial cells (24). Furthermore, it was shown that the T3SS-1 translocon members, SipB, SipC, and SipD, and presumably the assembly of the translocon, are essential for association with cultured mammalian cells. SipD is exposed on the bacterial surface and it may be localized at the tip of the needle complex. This potential position could then mediate intimate attachment. Next, upon contact with host cells, SipB and SipC may also become extracellularly exposed to contribute to this interaction (25).

1.2.3. Invasion of epithelial cells: cytoskeletal remodeling and induction of inflammation

Although SPI-1 independent mechanisms were discovered in the last years, the prevailing paradigm of *Salmonella* pathogenesis asserts that T3SS-1 is essential for bacterial invasion into

intestinal epithelium. The T3SS-1 injects effector proteins into the host cell cytosol, thus promoting actin cytoskeletal rearrangements and *Salmonella* engulfment (26). SPI-1 encodes T3SS-1, as well as regulators, effector proteins and chaperone proteins. More specifically, SPI-1 includes (Fig. 1):

- *prg*/*org* and *inv*/*spa* operons, encoding the needle complex,
- *sprB*, *hilC*, *hilD*, *hilA* and *invF* genes, encoding transcriptional regulators,
- *avrA, sptP and sipA* genes, encoding the effector proteins (other effectors are encoded elsewhere on the chromosome),
- *SipBCD*, encoding for the translocon, a pore-forming structure that embeds in the host cell membrane and delivers the effectors to the host cytosol,
- *sicP*, *sicA* and *invB*, encoding for chaperones which protect SPI-1-related proteins from degradation, prevent premature interactions, and/or mediate their recognition by T3SS-1.

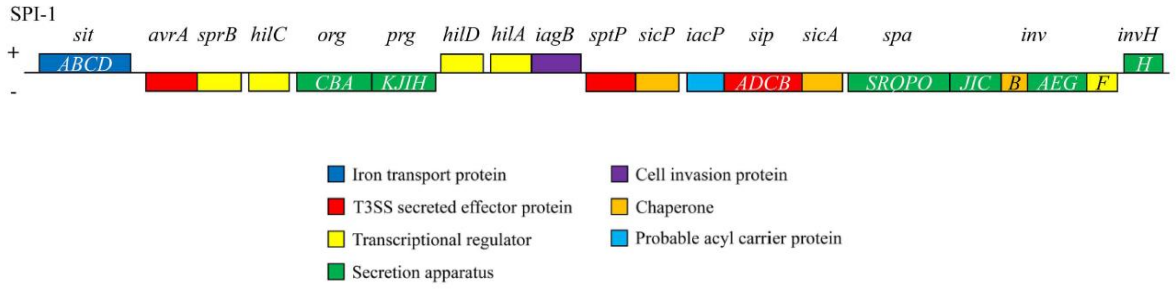


Figure 1: Schematic illustration of SPI-1: functional genetic categories are differentially colored (modified from 27).

Regulation of SPI-1, shown in Fig. 2, is extremely complex and depends on the balance of interactions among regulators encoded both inside and outside the island (28). SPI-1-encoded HilA directly activates the expression of the *prg/org* and the *inv/spa* operons, the latter encoding the AraC-like regulator InvF. InvF, in complex with SicA, activates the expression of a number of genes encoding secreted effectors including the sic/sip operon, and, outside SPI-1, *sopE* and *sopB*. The

three other AraC-like regulators, HilD, HilC, and RtsA, control expression of *hilA*. Each of these regulators is independently capable of inducing expression of *hilD*, *hilC*, and *rtsA* genes, as well as *hilA*, forming a complex coupled feed-forward loop to control SPI-1 expression (29). HilD is the dominant regulator of the system, as there is no *hilA* expression in its absence (30), and it is the primary site for SPI-1 regulatory signals. When the activating signals are sufficiently strong, HilD is expressed at a high enough level to overcome repression, activate the expression of HilC, RtsA, and HilA, and also further induce its own expression. The system is controlled primarily by affecting the threshold of HilD required for autoactivation (30). HilC and RtsA work to amplify and accelerate SPI-1 gene expression (31). In vivo, loss of either HilC or RtsA does not significantly attenuate intestinal invasion, whereas loss of both does (29), indicating that the amplification or acceleration provided by these loops plays an important physiological role.

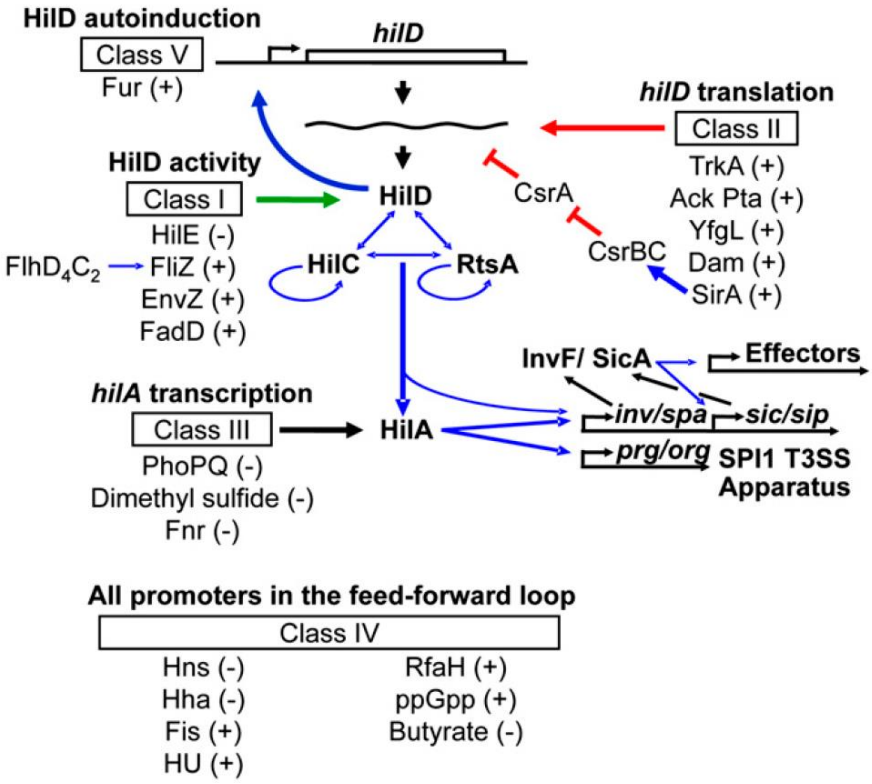


Figure 2: Working model for SPI-1 regulation. Blue lines indicate transcriptional regulation. Red lines indicate post-

transcriptional regulation. Green lines represent post-translational regulation. The effect of each regulator, positive (+) or negative (-) on *hilA* expression is indicated (28).

As the loops involving HilC and RtsA only additively contribute to the response, it was speculated that linking together multiple regulators could provide a robust and facile solution to strictly regulate SPI-1 expression, without commensurate effects on the underlying genes in case of changes in the regulatory architecture of the circuit (31).

A substantial number of genes and environmental conditions are implicated in SPI-1 regulation, which presumably ensures that SPI-1 is only expressed at the appropriate time and place within the host. Golubeva et al (28) determined where any given factor feeds into the SPI-1 regulatory circuit and grouped all of them into distinct classes (Fig. 2):

- Class I: regulation via the post-translational control of HilD. This class affects SPI-1 via control of HilD protein activity and includes the SPI-1 encoded negative regulator HilE and the activators FliZ (flagellar protein), EnvZ (two-component sensor kinase), and FadD (acyl-CoA synthetase).
- Class II: control of *hilD* mRNA stability or translation. The BarA/SirA two component system prevents block of *hilD* mRNA translation. The activators Dam, YfgL, Ack Pta, and TrkA belong to this class but their mechanism of regulation remains to be determined.
- Class III: regulation at the level of the *hilA* promoter. In this class there are factors that act at the level of *hilA* providing a potentially fast SPI-1 turn-off mechanism. The two component system PhoPQ, the global regulator Fnr and addition of dimethyl sulfide (found in the large intestine of mammals) to the growth medium belong to this group.
- Class IV: regulation of all SPI-1 regulatory promoters. This class is composed of a number of regulators and environmental conditions. The small nucleoid proteins H-NS and Hha

have been shown to directly bind to the promoter regions and silence transcription of SPI-1 genes including *hilD*, *hilC*, *rtsA*, and *hilA*. The presence of butyrate has the same effect. By contrast, the nucleoid proteins Fis, HU and RfaH, "temperature increase" and ppGpp positively control both *hilA* and *hilD* transcription.

- Class V: regulation by Fur. The global transcriptional regulator Fur is placed in a separated class because it seems to require both the HilD protein and *hilD* promoter to regulate *hilA*. Fur acts directly at the *hilD* promoter, but it is also proposed that it activates SPI-1 by reducing the H-NS silencing of *hilD* promoter, thus lowering the threshold of hilD required to activate the *hilD* expression.

All these regulators detect the best environmental conditions to allow invasion. SPI-1 is maximally transcribed during *in vivo* invasion of epithelial cells (32), whereas only those conditions resembling the host intestinal lumen, i.e., low oxygen and high osmolarity (33), and growth at the early stationary phase (34), activate these genes during *in vitro* assays.

It was observed that a *Salmonella* isogenic population (i.e. a population with identical genetic composition) can split into subpopulations with bistable expression of T3SS-1 (35). One subpopulation expresses T3SS-1 ($T1^+$ cells) and a fraction of those cells invades host tissue and evokes an inflammatory response that is beneficial for the *Salmonella* cells that do not express T3SS-1 ($T1^-$ cells). The T3SS-1 expression causes a slower growth of $T1^+$ cells, that has been associated with tolerance to antibiotics. Formation of a large $T1^-$ subpopulation may be crucial for intestinal persistence and to limit the generation of fast-growing avirulent mutants. SPI-1 bistability is thus a case of "cooperative virulence", that contributes to preservation of virulence in the *Salmonella* population.

Effectors involved in invasion, both encoded inside and outside SPI-1, have diverse functions in cytoskeletal remodeling and induction of the pro-inflammatory response (reviewed in 3,14,26, Fig. 3). SopB (SPI-5 encoded) is an inositol phosphatase that activates RHO GTPase-dependent actin rearrangements at the membrane of the host cell and recruits annexin A2, which functions as a platform for actin rearrangements. SopB also decreases levels of acidic lipids on the SCV, which may alter RAB family GTPase trafficking and antagonize SCV fusion with lysosomes. Furthermore, SopB influences the ability of the bacterium to transform follicular-associated epithelial cells into M cells through activation of WNT–β-catenin signaling and, in turn, receptor activator of nuclear factor-κB (NF-κB) ligand. Internalization of the bacterium is also promoted by SipA and SipC (SPI-1 encoded), which bind directly to actin at the site of insertion of the T3SS translocon (of which SipC is a component) and increase actin bundling at the site of bacterial entry into epithelial cells. SopE and SopE2 (encoded respectively in prophage SopEΦ and in the vicinity of a phage remnant) are guanidine exchange factors (GEFs) that activates Cdc42 and Rac-1, two Rho GTPases involved in the actin cytoskeletal rearrangements which induce localized actin polymerization and membrane ruffling.

The effectors described, together with SopA and SopD (encoded outside SPI-1), contribute also to intestinal inflammation by stimulating production of the pro-inflammatory cytokine interleukin-8 (IL-8) by epithelial cells through the mitogen-activated protein kinase (MAPK) and NF-κB pathways. It causes tight junction destabilization and stimulation of neutrophil transepithelial migration into the intestinal lumen. In addition to effector proteins, other virulence determinants like flagella, curli fimbriae and lipopolysaccharide (LPS) contribute to elicit inflammation.

Salmonella can cross the epithelial barrier also using a T3SS-1-independent mechanism. It was described a flagellum-dependent pathway for Peyer's patches invasion that requires the methyl-

accepting chemotaxis protein (MCP) Tsr and energy taxis toward host-derived nitrate. It is generated by inducible nitric oxide synthase (iNOS) in the ileal mucosa prior to infection (36).

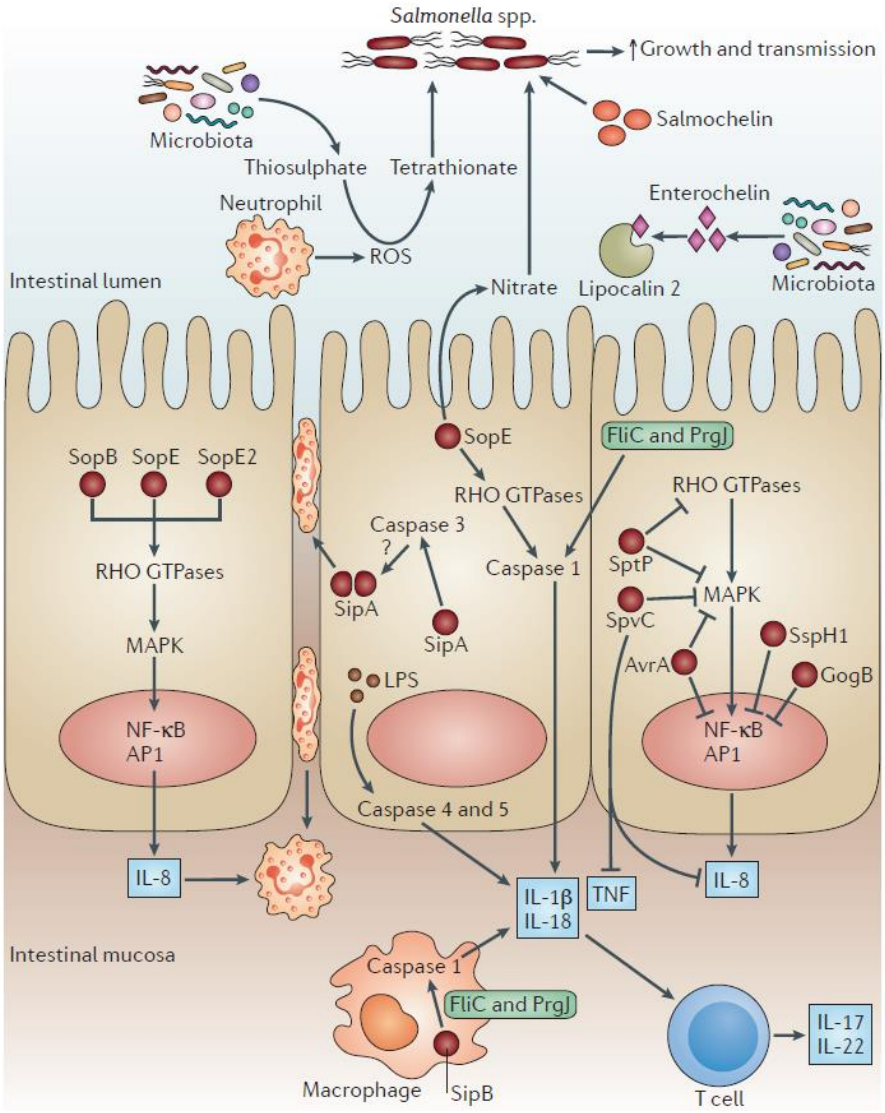


Figure 3: *Salmonella* interaction with epithelial cells and macrophages (14).

1.2.4. Downregulation of inflammation

After bacterial entry into host cells, cytoskeleton rearrangements and pro-inflammatory response are reversed. In fact, if on the one hand these events lead to epithelial damage (37), allowing

essential nutrients to become available for *Salmonella*, on the other hand their over-activation results in significant alteration of the host cell homeostasis that may be detrimental to the ability of the bacteria to survive, replicate, and disseminate inside the host (3).

Several so-called anti-inflammatory effectors downregulate inflammation by targeting MAPK or NF-κB pathways. The GTPase-activating protein SptP deactivates RAC1 and CDC42, thus reversing MAPK-mediated inflammation and IL-8 secretion (38). SPI-1 AvrA, SPI-2 SspH1, plasmid- encoded SpvC and GogB are also known to be implicated in downregulating inflammation (8).

1.2.5. Competition with the gut microbiota

A healthy gut microbiota provides 'colonization resistance' to pathogens like *Salmonella* by competing for nutrients (reviewed in 39). For example, gut bacteria ferment dietary fiber to short-chain fatty acids, which downregulate the expression of *Salmonella* invasion genes. Furthermore, some bacteria closely related to *Salmonella*, such as the probiotic *Escherichia coli* Nissle 1917, are able to outcompete *Salmonella* by limiting the access of the pathogen to iron, thus reducing *Salmonella* colonization. At the same time, *Salmonella* has several mechanisms to compete with the gut microbiota. Upon intestinal inflammation, lipocalin 2 accumulates in the lumen: this antimicrobial protein sequesters the bacterial siderophore enterochelin, required by enteric commensals for iron acquisition. By contrast, the *Salmonella* siderophore salmochelin is not sequestered by lipocalin 2, thereby providing to the pathogen a growth advantage over other enteric bacteria. *Salmonella* invasion results in the release into the intestinal lumen of reactive oxygen species (ROS), which react with thiosulphate, a respiratory by-product that is generated by the microbiota in the intestinal lumen, to generate tetrathionate. Tetrathionate is used by *Salmonella*, but not by the microbiota, as a terminal electron acceptor to support anaerobic or microaerophilic growth. Another mucosal response to *Salmonella* infection is the generatation of nitrate. Its production is due to the *Salmonella* effector protein SopE, which activates the

expression of nitric oxide synthase (iNOS) by a variety of cells, including intestinal epithelial cells, macrophages and neutrophils. iNOS converts L-arginine to nitric oxide (NO), a reactive nitrogen species (RNS) that inhibits several enzymes central to bacterial metabolism, except for *Salmonella* and other members of the Enterobacteriaceae family which can use nitrate in anaerobic nitrate respiration.

1.2.6. Survival and replication within SCV

Cytoskeletal rearrangements caused by *Salmonella* effectors disrupt the normal epithelial brush border and induce the subsequent formation of membrane ruffles that engulf adherent bacteria in SCVs. SPI-2 is required for survival and replication of *Salmonella* inside SCV. It encodes for a T3SS (T3SS-2), its traslocon machinery, and for multiple effectors injected into the host cytoplasm by means of T3SS-2. In general terms, SPI-2 harbors four types of genes: *ssa*, the genes encoding the T3SS-2 apparatus; *ssr*, encoding the regulators; *ssc*, encoding the chaperones; and *sse*, encoding the effectors (40, Fig. 4). Internalization of bacteria inside the SCV is followed by processes of SCV maturation and trafficking (3,14). At this stage the two effectors SopB and SsaB (which is also a component of T3SS-2) interact with this vesicular trafficking to escape from the normal degradation pathway, which ends upon fusion with lysosomes.

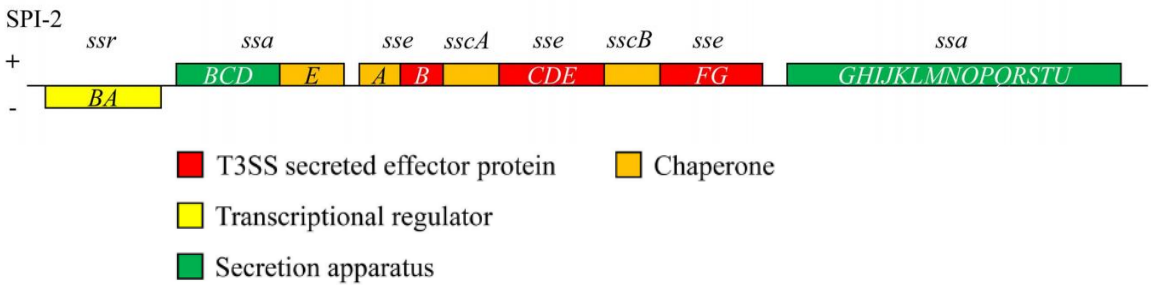


Figure 4: Schematic illustration of SPI-2: functional genetic categories are differentially colored (modified from 27).

SopB is also thought to mediate the formation of spacious phagosomes, which may provide a favorable environment where *Salmonella* can reside and build its replicative niche (41). Moreover, SifA, a major SPI-2 virulence protein that is localized in the SCV membrane, has also been reported to interact with the endocytic pathway (42).

During SCV maturation, *Salmonella* induces *de novo* formation of an F-actin meshwork around bacterial vacuoles, a process which is termed vacuole-associated actin polymerization (VAP) and is required for maintenance and stability of the SCV membrane. Recent experiments have revealed that not only the classical T3SS-2-dependent effectors (e.g., SspH2, SseI, and SpvB) are involved in this process. The T3SS-1 effector and actin binding protein SipA, which has been reported to persist after bacterial internalization, is exposed on the SCV and stabilizes the actin filaments produced during VAP (43). After maturation and actin polymerization events, SCV migrates toward a perinuclear position depending on the two microtubule proteins kinesin and dynein, which transport cargo inside cells toward the cell periphery and to the nucleus, respectively (44). SCVs migrate to a perinuclear position presumably to facilitate interception of endocytic and exocytic transport vesicles to obtain nutrients and/or membrane fragments. This event appears to be essential for bacterial replication. In addition, it has been observed that intracellular *Salmonella* can induce the formation of long filamentous membrane structures called *Salmonella*-induced filaments (SIFs) through several effectors, such as SifA, SipA, SseJ, SseF, and SseG (reviewed in 3,14).

Salmonella exhibits contrasting lifestyles inside host cells, with different replication rates and subcellular locations. *Salmonella* population inside SCV can be divided in two subpopulations, one that rapidly replicates and the other that does not replicate but are still viable (45). These non-replicating variants are recalcitrant to be killed by antibiotic (termed antibiotic tolerance) and are

known as persisters. *Salmonella* can also replicate in the cytosol of epithelial cells to higher numbers than vacuolar bacteria, a phenotype dubbed “hyper-replication" (46). The two intracellular populations of bacteria are transcriptionally distinct: the intravacuolar bacteria are SPI2-induced whereas the cytosolic bacteria are SPI1-induced and flagellated. Epithelial cells containing hyper-replicating SPI1-induced *Salmonella* undergo inflammatory cell death, marked by loss of plasma membrane integrity and activation of caspase-11. Ultimately these cells are extruded from monolayers and released into the extracellular milieu.

1.2.7. Programmed cell death and systemic dissemination

Internalization of the infecting *Salmonella* within the SCV is followed by induction of host cell death or systemic spread through other target organs, primarily liver and spleen.

Host cell death is a host strategy to inhibit intracellular replication of bacteria, but can also be exploited by *Salmonella* for further dissemination. Epithelial cell death can be induced maintaining an intact plasma membrane to prevent release of inflammatory intracellular contents. The apoptotic bodies can be taken up by phagocytes allowing for degradation in a generally non-inflammatory process. The event is induced by caspase-3 and T3SS-2 after prolonged exposure, at least 12h after *in vitro* infection (47). The main *Salmonella* effectors that promote epithelial cell death are SpvB and SlrP (reviewed in 3). SopB, AvrA and flagellin are reported to counteract early induction of apoptosis.

Epithelial cells and macrophages can undergo pyroptosis, an inflammatory form of cell death triggered by *Salmonella* through different mechanism, including secretion of flagellin and SipB into host cell cytosol by T3SS-1, which in turn activates the NLRC4 inflammasome (reviewed in 48, 49). Other mechanisms include activation of the NLRP3 inflammasome by an unknown ligand, and activation of caspase-11 by cytosolic LPS via unknown receptors. This cytotoxicity in macrophages can appear very rapidly after phagocytosis (requiring T3SS-1) or can be induced several hours later

(essentially requiring T3SS-2). *Salmonella* also exploits the induction of necroptosis in macrophages, a form of cell death dependent on the induction of type I interferon (IFN) signaling by *Salmonella*, which in turn drives the activation of the RIP1 and RIP3 kinases.

Not all macrophages that engulf *Salmonella* die as a result of infection, and such cells constitute a niche that enhances *Salmonella* persistence during chronic infection. *Salmonella* was found to preferentially survive in macrophages with an anti-inflammatory M2 phenotype (48). A subset of these macrophages in the spleen is characterized by having engulfed leukocytes, including B and T cells, and are known as hemophagocytic macrophages. They are defective in limiting the replication of *Salmonella*, likely because they secrete very low levels of pro-inflammatory cytokines. Another factor that promotes *Salmonella* replication within M2 macrophages is the altered metabolism of these phagocytes. During chronic *Salmonella* infection, upregulation of peroxisome proliferator-activated-receptors (PPARs) in M2 macrophages results in increase of glucose availability to the pathogen, thereby enhancing its replication inside these cells and sustaining its long-term survival within the host. Another means by which *Salmonella* establishes chronic infection in the liver is by direct inhibition of CD4+ and CD8+ T cell proliferation. It was shown to be dependent on the secretion of L-asparaginase II in a mechanism still not known.

1.3. Host adaptation

Salmonella serovars vary widely in their capacity to cause disease in different host species. Three groups of *Salmonella* serotypes can be defined based on their host-tropism:

1. Host unrestricted (Generalists) serovars, which generically cause infections in a number of unrelated hosts,

2. Host restricted (Specialists) serovars, which are almost exclusively associated with a certain host

species and can cause serious systemic infection within that host,

3. Host adapted serovars, which are prevalent in one particular host species but can also cause disease in other host species.

S. Typhimurium and *S*. Enteritidis are considered host unrestricted serovars. Interestingly, they cause different diseases in different animal species (50). In calves, serovar Typhimurium (and rarely Enteritidis) causes enterocolitis, and animals might succumb to dehydration. In newly hatched chicks, these serovars cause systemic disease and diarrhea, whereas older chickens are asymptomatic carriers. In immunocompetent humans, serovars Enteritidis and Typhimurium cause localized self-limiting enterocolitis while systemic disease may develop in immunocompromised individuals. Finally, serovars Enteritidis and Typhimurium cause a systemic typhoid fever-like disease in susceptible mouse strains. Typhi, Paratyphi A/B/C, Gallinarum, and Pullorum are examples of host-restricted serovars (51,52). In each case, within-host evolution has selected for variants that evade detection by the host innate immune system, with concomitant blunting of the inflammatory response in the intestinal mucosa and dissemination to systemic sites. In fact, *S*. Typhi and *S*. Paratyphi A cause a systemic typhoid disease in humans, serovar Pullorum causes the systemic pullorum disease in poultry, with high mortality in freshly hatched chicks, serovar Gallinarum causes the severe systemic fowl typhoid disease in poultry and a few other avian species. By contrast, experimental evidences suggest that almost none of these serovars are able to cause disease in the mouse model of *Salmonella* infection. *S.* Dublin is a serovar adapted to cattle, but it can also infect and cause disease in different hosts, including humans. In laboratory settings serovar Dublin was found capable of causing typhoid fever-like infections in mice (50), even though at lower extent than *S*. Typhimurium and *S*. Enteritidis.

Epidemiological evidence supports various levels of host adaptation even among strains of broad host range serovars. For example, *S*. Typhimurium has been considered the prototypical broad-

host-range serovar, however, the epidemiology of *S*. Typhimurium suggests a more complex mixture of 7 so-called pathovariants, some of which highly host adapted. For example, particular phage types such as DT2 or DT99 can cause systemic infections in pigeons and the multi-locus sequence type ST313 causes systemic infection in humans (53, Fig. 5). Thus, it appears that both inter- and intra-serovar variations have a role in host range and disease severity. Emergence of host-adapted pathovariants highlights that adaptation to new hosts is an ongoing process.

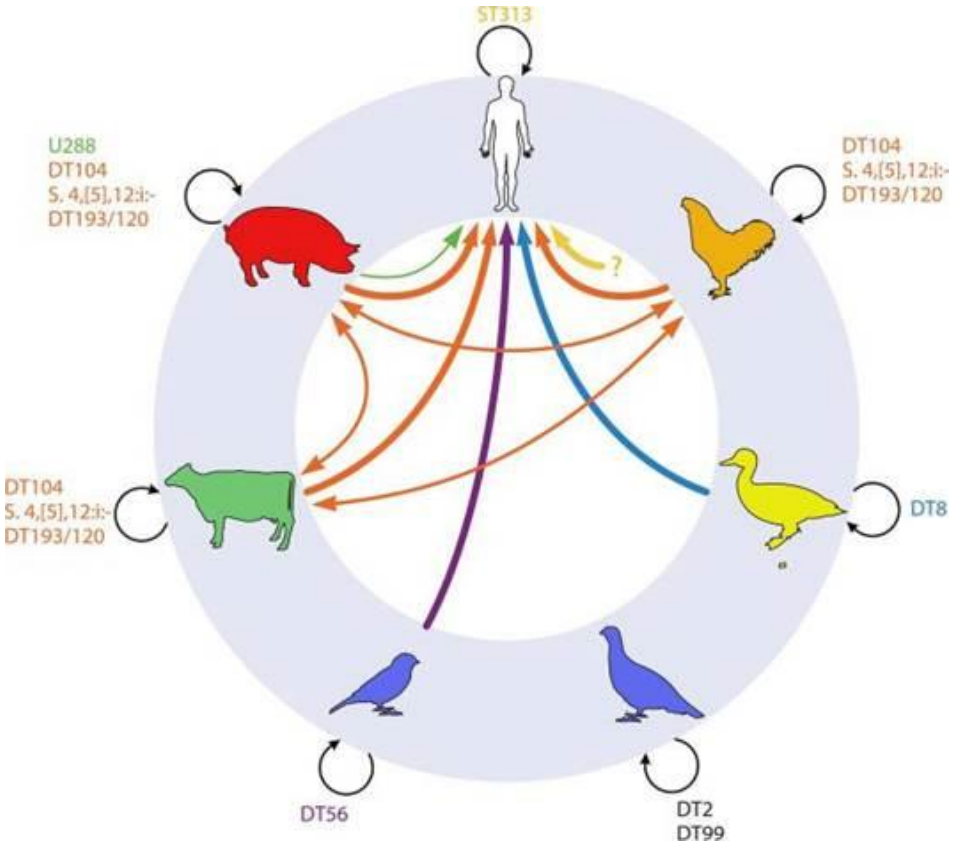


Figure 5: Host range and plausible transmission routes of major *S*. Typhimurium pathovariants in humans, pigs, cattle, wild avian species, ducks, and poultry (53)

1.3.1. Genetic features correlated to host adaptation

Understanding the basis of host adaptation could be critically important to better comprehend molecular pathogenesis and the evolution of pathogenic microbes, and to assess the risk of

pathogens to cross the species barrier and infect new hosts. Even if the overall mechanism of host specificity remains still largely elusive, it is known to be associated with a number of genomic signatures, including genomic decay, generation of different allelic variants by point mutations, genomic rearrangements, and gene acquisition by lateral gene transfer (reviewed in 51, Fig 6).

The genome of host-restricted serovars commonly shows signs of genomic decay, including gene deletion and gene inactivation through point mutation (pseudogene formation). In *S*. Typhi and *S.* Paratyphi A, >4% of the genome-coding sequences are pseudogenes, compared to 0.9% in *S.* Typhimurium. In *S*. Gallinarum, the proportion of pseudogene is >7%. It was speculated that host-restricted serovars have lost genes no more necessary for their new life style, and therefore no more under selection. Specialists are associated with septicemic infections in one species. Accordingly, many pseudogenes encoded functions necessary for intestinal colonization and virulence in secondary hosts. It was shown that *S.* Gallinarum has lost genes from several metabolic pathways, like genes for the catabolism of maltodextrins, (*D*)-glucarate and 1,2-propanediol, and a number of T3SS effector proteins. This serovar is also nonmotile, because carries mutations in five genes required for chemotaxis and motility (*cheM*, *flhA*, *flhB*, *flgK* and *flgI*) (54).

Instead of being lost, genes may also evolve differentially in response to different hosts, accumulating single nucleotide polymorphisms (SNPs). Eswarappa et al. (55) identified T3SS translocon components of both SPI-1 (SipD) and SPI-2 (SseC and SseD) which have evolved differently in different serovars of *Salmonella*. SipD is conserved between *S*. Typhi and *S*. Paratyphi, but the nonsynonymous SNPs distances (DNs) calculated between the human-restricted serotypes and other serotypes like *S*. Typhimurium, *S*. Choleraesuis, and *S*. Enteritidis are large. The entry defect of *S*. Typhi deleted for *sipD* (Δ*sipD*) in epithelial cells was abolished when the *sipD* allele of S. Typhi is expressed but not after complementation with the *sipD* allele of *S*.

Typhimurium. However, the entry defect of *S*. Typhimurium Δ*sipD* is abolished when the *sipD* allele of either *S*. Typhi or *S*. Typhimurium is expressed. Human restricted serovars thus appear to have evolved a *sipD* allelic variant necessary for its specific mechanism of invasion. DN values of *sseC* between *S*. Typhi and *S*. Paratyphi and between *S*. Typhimurium and *S*. Enteritidis were significantly smaller than the other combinations of serovars, suggesting that *sseC* has evolved differently in human-restricted serovars and in serovars that can infect multiple hosts.

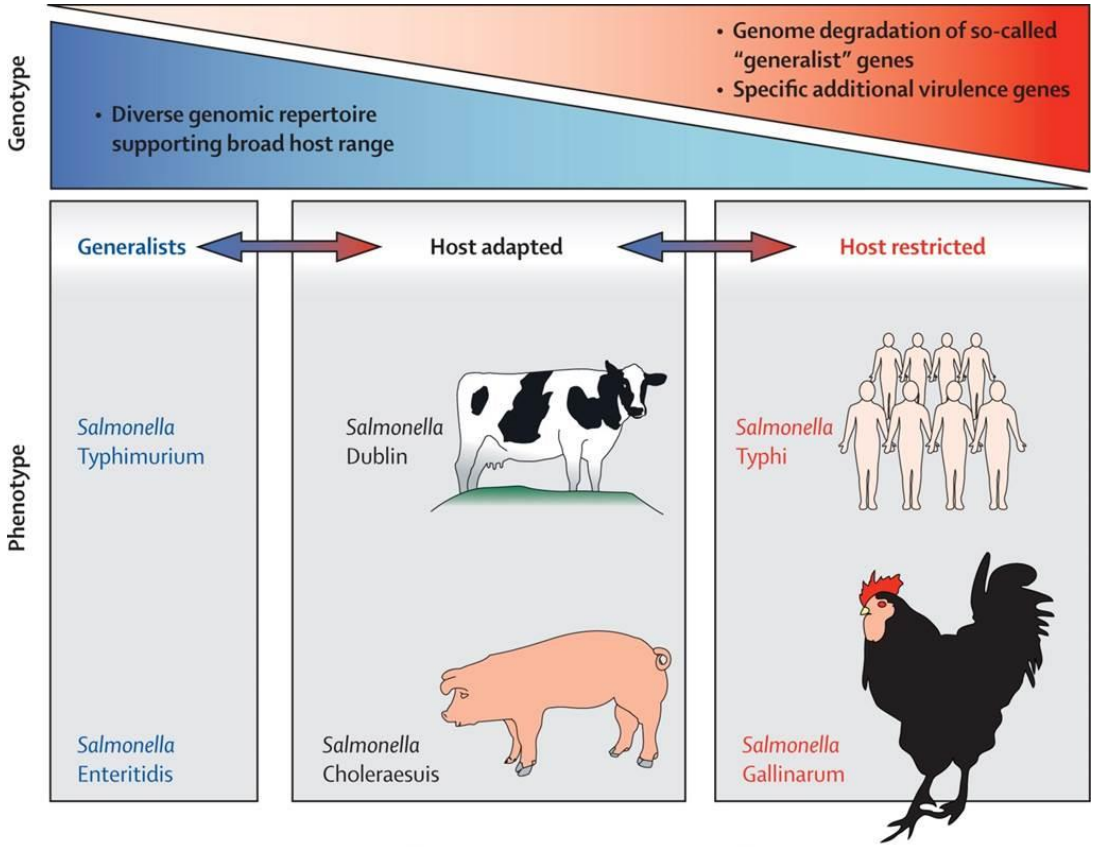


Figure 6: Genetic Features of host adaptation in *Salmonella* (modified from 57).

Min Yue et al (56) illustrated population and functional analyses of host-associated non-synonymous (ns) SNPs in *fimH*, encoding for the type 1 fimbrial adhesin, highlighting the role of key allelic residues in host-specific adherence *in vitro*. The phylogeny made on 152 unique *fimH*

sequences from 76 different serovars demonstrated a strong serovar-specific lineage, with strains of the same serovar clustering either together or in a few distinct branches. To evaluate the biological relevance of the *in silico* determined associations, they assessed the host-specific binding properties of 16 *fimH* allelic variants from major serovars by expressing them in the context of *Salmonella* type 1 fimbriae on recombinant *E. coli*. Binding assays using human, bovine and porcine intestinal epithelial cells, and hepato-epithelial chicken cells, revealed that several allelic FimH proteins conferred significant host-specific binding.

Genomic rearrangements are another genomic signature associated with specialists but absent in generalists (51). Large rearrangements are commonly associated with a lower growth rate, because they may alter gene dosage, replication-transcription conflicts, or chromosome symmetry. While *S*. Typhimurium needs to multiply at a maximum growth rate in the intestine because successful transmission requires bacterial numbers in the feces to reach a sufficient threshold (58), the transmission success of host-restricted *Salmonella* serovars is aided by chronic carriage, which is associated with slow growth. Therefore, it was speculated that selective constraints that prevent chromosomal rearrangements may no longer be applied to host-restricted *Salmonella* serovars. Genome rearrangements were found for example in *S*. Typhi (59), *S*. Paratyphi (60) and *S*. Gallinarum (61).

A third genomic signature associated to host adaptation is the acquisition of genetic material by horizontal gene transfer (HGT), found in genomes both of specialists and generalists. HGT is mediated by a number of mechanisms including conjugation, transformation, phage-mediated transduction and lysogeny. Conjugation is the dominant mechanism for integrative conjugative element (ICEs, like SPIs) and plasmids, but likely plays an important role in transfer of other mobile

genetic elements (MGEs), such as transposons. One example of acquisition of genes implicated in host adaptation was described in *S*. Typhi. It horizontally acquired SPI-7, which includes the regulator *tviA* and the capsular polysaccharide genes encoding for the Vi Capsular Antigen. *S*. Typhi, activating the expression of SPI-7 genes, evades innate immunity once entered in the intestinal mucosa, because the Vi Antigen obstructs bacterial-guided neutrophil chemotaxis (62). SPI-7 is present even in the host-adapted serotype Dublin and the host-restricted *S*. Paratyphi C. Both serotypes, like *S*. Typhi, cause systemic infections in their prevalent or unique or hosts, suggesting that this mechanism of immune evasion could be critically important for their pathogenesis (63).

1.3.2. Host-adaptation in terms of evolution

The predominant isolation of one serovar from a particular host over another serovar could be the result of competitive exclusion (64). The competitive exclusion is a process by which an organism prevents another one from colonizing a given environment because it is better able to establish and maintain itself in that environment. By considering the selective pressure imposed by the host immune response on pathogen populations, it is possible that *Salmonella* serotypes associated with a specific host may represent the winners of between-serotypes competition. Host adaptation may thus result in a natural balance in which one host-adapted *Salmonella* serotype competitively excludes other members of the same serogroup from circulating in a particular host (65). How does specificity evolve? Host–pathogen specificity could arise as a result of evolutionary constraints, when the traits that help a parasite to exploit one host make it less able to attack other hosts (66). Observations of within-host evolution of *Salmonella*, that led disseminated disease, indicate that it is normally associated with a decrease in host range, because mutations benefitting replication in systemic sites in one host species do not result in the same benefit in other host species (52). There is a trade-off for the relative benefit to transmission in a single host

species population, and the decrease in host range due to loss of fitness in other hosts. For example, in the case of *S.* Typhi, host restriction to humans was accompanied by a distinct transmission strategy involving dissemination to systemic sites to gain access to the gall bladder, establishing long-term persistence. The decrease in host-range due to host restriction was counterbalanced by the increased longevity of transmission (52).

1.4. *Salmonella* Derby

1.4.1. Epidemiology and Phylogeny

Host associations can be inferred from isolation statistics compiled during background surveillance programs, where a particular serotype or subtype is consistently isolated from the same sub-set of host species, spanning years and geographical areas. Enter-net is the network of national and regional laboratories dedicated to tracking foodborne infections in Europe. Each laboratory utilizes standardized genotyping methods and shares information in real-time. The surveillance system is coordinated by the European Center for Disease Prevention and Control (ECDC). The latest report (67) on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks produced by EDCD and The European Food Safety Authority (EFSA) indicates that serovar Derby is the fifth most frequently reported serovar among human cases of infections within EU (0.9%), after *S*. Enteritidis (59%), *S*. Typhimurium (13.6%), *S*. Typhimurium monophasic variants (including 'monophasic variant of S. Typhimurium', '4,[5],12:i:-' and '4,,12:i:-') (4.7%) and *S*. Infantis (2.3%). *S*. Derby, which is seventh in the ranking of veterinary isolates (3.3%), is isolated mainly from swine (64.4%), but also from turkey (21.0%), broiler (11.3%) and bovine sources (3.3%). These descriptive results tend to identify pigs as the main animal reservoir for *S*. Derby, followed by poultry. In pigs, *S*. Derby is the third ranked serovar (19.2%), after *S*. Typhimurium monophasic variants (34.1%) and

S. Typhimurium (29.5%).

S. Derby is a polyphyletic serotype. Four MLST profiles, ST39, ST40, ST71 and ST682, have been described in different works (68, 69, 70, 71, 72, 73). Sévellec et al (73) identified four clusters by SNPs-based phylogenic analysis of 140 genomes of *S*. Derby isolates. These groups are fully consistent with the ones identified by MLST analysis (Fig. 7). The four lineages are mainly associated to different hosts: in the already cited works, it was described that ST40, ST39, and ST682 are associated with pork, while ST71 is associated with poultry. Genetic differences among these lineages were analyzed. SPI-1, -2, -3, -4, -5 were detected within *S*. Derby isolates of all sequence types, with an average sequence identity of 97%. Lineage-specific differences were observed in the genes *sseB*, *sseC*, and *sseD* coding for the SPI-2 translocon.

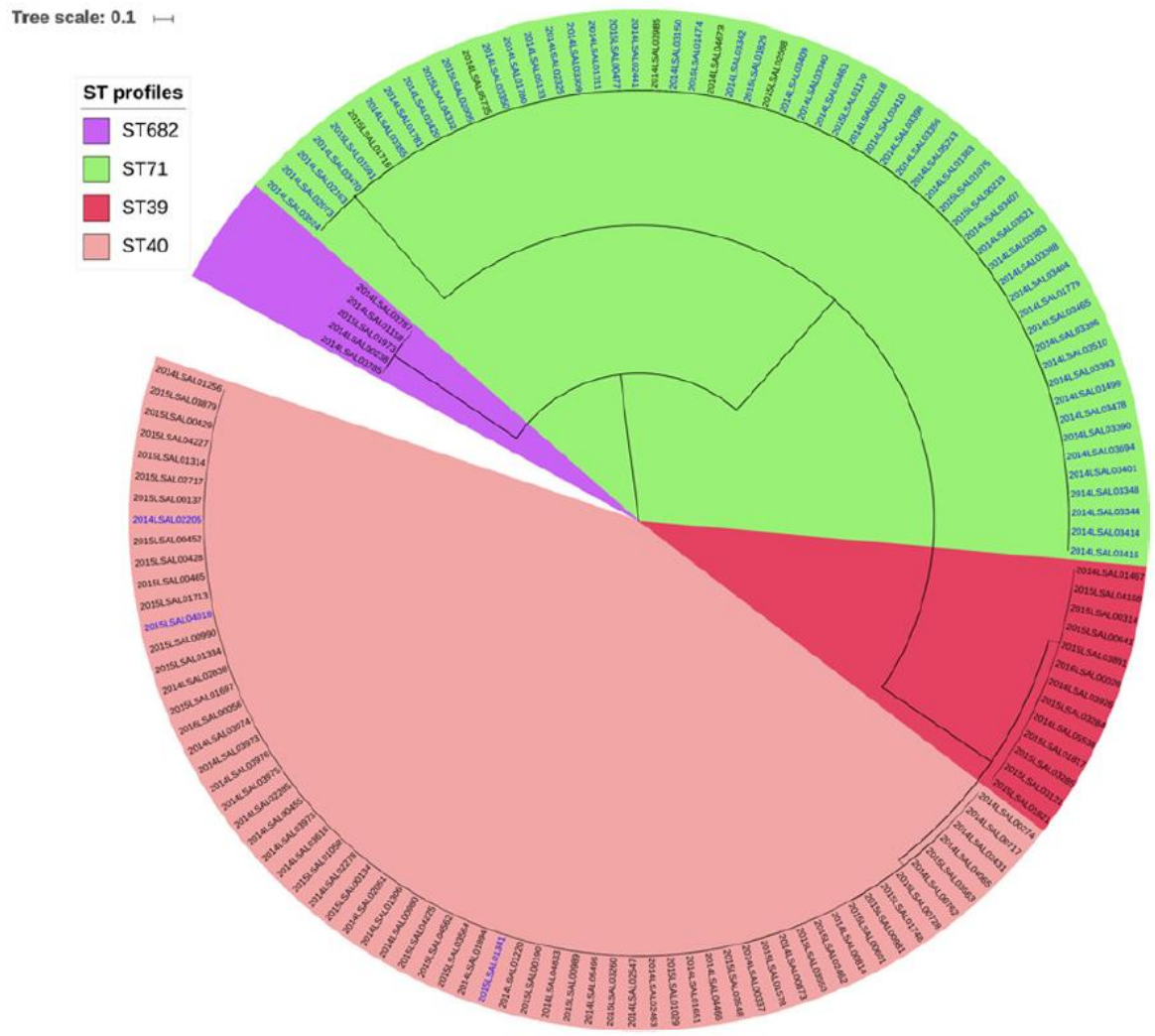


Figure 7: Phylogenetic SNP tree for 140 strains of S. Derby constructed under the maximum likelihood criterion using the GTR-gamma model. The scale bar indicates the number of substitutions per site. Four clusters were identified and named ST39, ST40, ST71, and ST682 with reference to the MLST profiles. The labels in blue correspond to the strains of the poultry sector. The black labels correspond to the strains of the pork sector (73).

SPI-3 was the most variable SPI within the genomes analyzed. A deletion of the left-terminal genes *rhuM* and *sugR* was observed: *rhuM* is present only in the lineage ST682, while *sugR* gene was found fragmented in all the lineages. ST40, ST39, and ST71 possess an integron in their SPI-3 containing seven genes related to the adhesion structures that were reported as characteristic of *S*. Derby (74). SPI-23, described by Hayward et al. as playing a role in adhesion and invasion of porcine tissues (75), is present in isolates belonging to lineages ST39 and ST40, related to pig, and absent in ST71, associated with poultry. It is also absent in the lineage ST682, associated to pig,

but it could be due by the fact that this lineage is the most genetically distant from ST39, ST40 and ST71, suggesting likely different evolutionary paths.

1.4.2. S. Derby epidemiological analysis in Emilia Romagna

The enteropathogens surveillance system in Italy is organized both at a national and regional level. In Emilia Romagna a network of 10 laboratories has been coordinated since 2012 by the regional Enter-net reference center located in Parma, included in the Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna. This regional Enter-net reference center collects every year 800-900 human and 700-800 veterinary isolates of *Salmonella*. Each isolate is subjected to serotyping and PFGE-based typing. A multiplex PCR-based method is also performed to discriminate between diphasic and monophasic *S*. Typhimurium.

The total number of sporadic isolates (isolates not belonging to identified outbreaks) collected from 2012 to February 2017 from the 4 main hosts (human, poultry, cattle and swine) are 6,335. The majority of veterinary *Salmonella*e were isolated from swine (1,164 compared to 624 from poultry and 174 from cattle), which is therefore the main *Salmonella* reservoir in this region. Table 1 shows, for the 10 more diffused serovars in Emilia Romagna, the number of isolates collected from the four main hosts and their proportion in each host (calculated on the total number of *Salmonella* isolates collected per host). The two main serovars isolated from swine are Typhimurium monophasic variant (36.94%) and Derby (27.06%). *S*. Typhimurium monophasic variant is also the most isolated serotype in humans (46.26%), while *S*. Derby isolates from this host are more the ten times less abundant (2,74%). These epidemiological data are in line with those reported by EFSA and ECDC (67): they indicate that *S.* Typhimurium monophasic variants is equally distributed in human and swine, whereas *S*. Derby main reservoir is swine.

Table 1: The table reports, for the 10 more diffused serovars in Emilia Romagna, the number of isolates collected from

the four main hosts (n°) and percentage of isolates of each serovar on the total number of isolates collected from each host.

	human		poultry		cattle		swine	
Serovar	**n°**	**Percentage**	**n°**	**Percentage**	**n°**	**Percentage**	**n°**	**Percentage**
Monophasic	2023	46,26%	17	2,72%	56	32,18%	430	36,94%
Typhimurium	610	13,95%	29	4,65%	49	28,16%	58	4,98%
Enteritidis	444	10,15%	52	8,33%	2	1,15%	11	0,95%
Napoli	148	3,38%	3	0,48%	1	0,57%	0	0,00%
Derby	120	2,74%	5	0,80%	6	3,45%	315	27,06%
Brandenburg	103	2,36%	0	0,00%	0	0,00%	37	3,18%
Infantis	99	2,26%	223	35,74%	6	3,45%	22	1,89%
Give	58	1,33%	4	0,64%	3	1,72%	21	1,80%
Panama	56	1,28%	0	0,00%	0	0,00%	4	0,34%
Rissen	43	0,98%	5	0,80%	5	2,87%	106	9,11%

PFGE typing identified 101 different *S*. Derby pulsotypes. Cluster analysis on pulsotypes was carried out based on the unweighted pair group method with arithmetic averages (UPGMA) to obtain a dendrogram of pulsotype relationships (Fig. 8). The two most prevalent PFGE profiles in swine are SXB_BS.0204 and SXB_BS.0056, representing 20% of the total S. Derby isolates from this host. These two profiles have different prevalences in human. SXB_BS.0204 is less isolated in human (1 human isolate, 0.9%) than in swine (27 swine isolates, 9.4%) while PFGE_B is proportionally isolated in human (15 human isolates, 13%) as well as in swine (26 swine isolates, 9.7%). These PFGE profiles show a high percentage of similarity (about 90%).

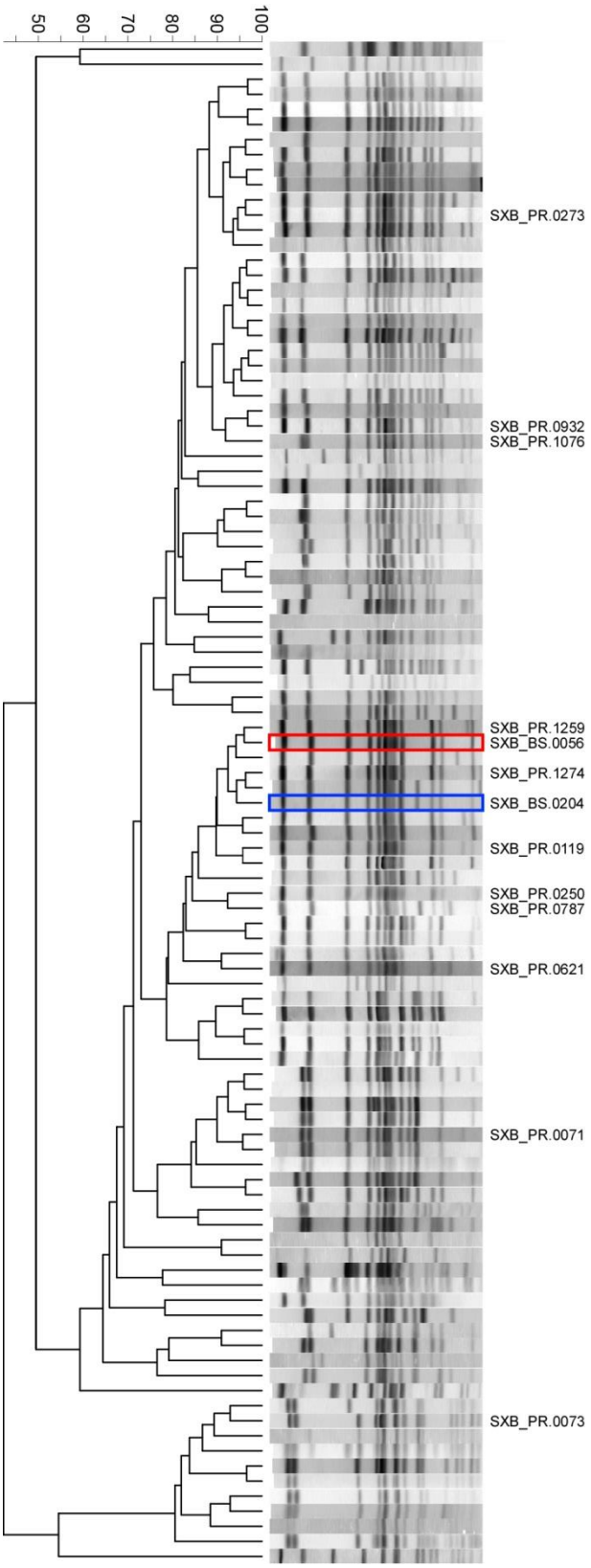


Figure 8: S. Derby PFGE dendrogram. Names assigned to the pulsotypes of isolates used in this work are indicated. The SXB_BS.0056 and SXB_BS.0204 PFGE profiles are surrounded in red and blue, respectively.

To quantify the differences in distribution of SXB_BS.0204 and SXB_BS.0056 profiles in human and swine the selection ratio (SR) of each PFGE profile was computed through multinomial simulations. The SR is defined as the proportional use of a resource divided by its proportional availability. In this context, it corresponds to the ratio between the fraction of swine isolates with a specific profile on all swine isolates and the fraction of human isolates with the same profile on all the human isolates, considering only *S*. Derby isolates. Where the SR=1 the fraction of swine isolates with a specific profile is directly proportional to the fraction of human isolates with the same profile so this profile represents opportunistic pathogens. An SR>1 [SR<1] means there are relatively more [less] isolates with a given PFGE profile from human than swine. Le multinomial simulations were used to evaluate if the difference in distribution in the two different host for each pulsotype is statistically significant. SXB_BS.0204 is significantly less isolated in human than in swine (SR= 0.0875, *p*-value<0.001), while SXB_BS.0056 is isolated in human as well as in swine (SR=1.4583, not significantly different than 1). The SR calculated on isolates belonging to all the other PFGE profiles combined is not significantly different from 1 (Fig. 8).

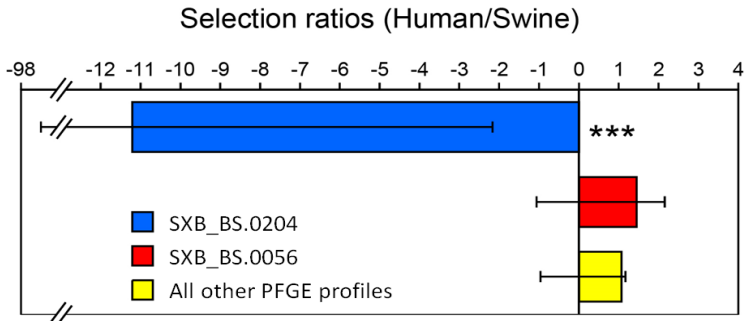


Figure 9: Selection ratios of the most prevalent PFGE profiles in swine, SXB_BS.0204 (blue) and SXB_BS.0056 (red), and all other S. Derby profiles combined (yellow). Negative values are calculated as (-1/SR).

1.5. Thesis outline

The last report on zoonoses published by EDCD and EFSA indicated that *Salmonella* is the second most diffused cause of foodborne disease in the European Union (67). Discovering the molecular determinants involved in host specificity would be critically important to assess the risk of *Salmonella* spread to different hosts and consequently its transmission along the food-chain. In Emilia Romagna, swine is the main *Salmonella* reservoir. The two main serovars isolated from swine are Typhimurium monophasic variants and Derby. *S*. Typhimurium monophasic is also the most isolated serotype in humans, whereas *S*. Derby isolates are only rarely found in humans. Intriguingly, different prevalence in humans were detected for the two most prevalent S. Derby PFGE profiles in swine: SXB_BS.0204 is about ten time less isolated in humans than SXB_BS.0056.

Based on these epidemiological data, the aim of this PhD project was to detect and characterize the genetic determinants responsible for the different distribution of SXB_BS.0204 and SXB_BS.0056 isolates in human.

Here, each step performed to answer the main question of the project is presented, describing the approach used to validate the hypothesis formulated in each step.

First of all, we evaluated whether the different prevalence in humans of SXB_BS.0204 and SXB_BS.0056 isolates was due to differences in virulence. Infection assays were performed on *C. elegans* and human and swine epithelial cell lines, already used as model for studying *Salmonella* pathogenesis. *In vitro* cell culture, but not *C. elegans* infection assays allowed to observe phenotypic differences in virulence between SXB_BS.0204 and SXB_BS.0056 isolates. Tested isolates were defined non-virulent and virulent based on their different infection efficiency, especially in human cells.

To assess the phylogenetic relationships between virulent and non-virulent isolates, whole genomes of 45 isolates were sequenced, to generate a Maximum Likelihood tree.

As host specificity is associated with a number of genomic signatures including genomic decay and gene acquisition by horizontal gene transfer, we hypothesized that virulent isolates carry exclusive genes responsible for their virulence phenotype. Thus, an orthology analysis was performed to find genes exclusive of virulent isolates. We obtained deletion mutants for the selected genes, which were tested for virulence in *in vitro* cell culture infection in comparison to the respective wild type strains.

As we found that the higher infection efficiency of virulent isolates was not due to the presence of exclusive genes, we hypothesized that allelic variants are involved in the different virulence phenotypes observed. SNPs were selected which discriminate virulent from non-virulent isolates. NV isolates carry two non-sense mutations in *ydiV* and *yhaK* genes, encoding respectively for a repressor of flagellar genes and for a protein of unknown function involved with oxidative stress. *ydiV* and *yhaK* virulent alleles were replaced in ER1175 (V isolate) with the non-virulent variants to assess their involvement in the attenuate virulence of NV isolates. Mutants did not show any differences in their virulence phenotype compared to the wild type strain. The analysis detected also two allelic variants of HilD, the major transcriptional activator of *Salmonella* Pathogenic Island (SPI) 1 and other virulence genes. SPI-1 encodes for the type three secretion system (T3SS), a ‘‘molecular syringe’’ that injects effector proteins into the host cell cytosol, allowing *Salmonella* to invade epithelial cells.

To assess the role of *hilD* allelic variants in the observed virulence phenotypes, the virulent allele of ER1175 was replaced by recombination with the non-virulent allele and vice versa in ER278 isolate (non-virulent). Both strains were also deleted for *hilD* and *invA* (encoding for a main structural component of the T3SS-1) to assess respectively if the deletion of the entire gene and the loss of a functional T3SS-1 confers the same virulence phenotype of the non-virulent *hilD* allele.

The observed phenotypes suggest that the non-virulent *hilD* allele produces a non-functional protein, and that loss of a functional HilD elicits the same virulence phenotype of loss of a functional T3SS-1. To confirm this hypothesis, differential expression analysis was performed by extracting RNA both in *in vitro* growth condition, to perform the RNA-seq, and during human cell line infection, to perform quantitative RT-PCRs on SPI-1 and invasion-related genes.

Lastly, the mouse model for *Salmonella* infection was used to assess possible differences in *in vivo* host-bacterial interaction caused by *hilD* allelic variants.

2. Materials and Methods

2.1. Bacterial strains

The Regional Reference Centre for Foodborne Diseases of Emilia Romagna, located in IZSLER-Parma, collected 435 *S*. Derby sporadic isolates from human and swine sources between January 2012 and February 2017. Each isolate was genotyped by Pulsed-Field Gel Electrophoresis following the PulseNet Standard Operating Procedure (76). All *S*. Derby strains selected for this work, their metadata and the assays to which they were subjected are listed in Table 2. All mutants generated are listed in Table 3. Bacteria were grown in Luria Bertani (LB) broth (5 g/L NaCl, 10 g/L tryptone, 5 g/L yeast extract), and/or on LB agar plates with 15 g/L agar. Super Optimal broth with Catabolite repression (SOC) medium was used for the recovery of transformants (77). The growth temperature used was 37°C except for the strains containing the temperature-sensitive plasmid pKD46 (78), which were grown at 30°C. Unless otherwise specified, bacteria were grown with shaking (200 rpm). Mutants were generated from *S*. Derby isolates ER278, ER1175 and N11. When needed, antibiotics were added at the following concentrations: 100 mg/ml ampicillin; 20 mg/ml chloramphenicol; 15 mg/ml gentamicin.

Table 2: *S*. Derby isolates used in this work. The table shows for each isolate: PFGE Genotype, date of receipt, isolation source, and the assays to which it was subjected (marked by 'x').

Strains	PFGE Genotype	Date of Receipt (DD/MM/YYYY)	Isolation matrix	*C. elegans infection assays*	*In vitro* cell culture infection assays	antibiotic resistance screening	Sequencing
AH35	SXB_BS.0056	18/06/2013	human		x	x	x
AL2	SXB_BS.0056	06/08/2013	human		x	x	x
BA77	SXB_BS.0056	07/08/2014	human		x	x	x
ER125	SXB_BS.0056	27/09/2013	swine-salami	x	x	x	x
ER1625	SXB_BS.0056	22/12/2015	swine-salami		x	x	x
AF47	SXB_BS.0056	18/04/2013	human		x	x	

ER1364	SXB_BS.0056	29/07/2015	swine - faeces		x	x	
ER1857	SXB_BS.0056	02/03/2016	swine-carcass		x	x	
ER1863	SXB_BS.0056	26/05/2014	swine-meat		x	x	
AH66	SXB_BS.0056	03/06/2013	swine-env*				x
AL3	SXB_BS.0056	06/08/2013	human				x
AP36	SXB_BS.0056	16/10/2013	human				x
AP8	SXB_BS.0056	10/10/2013	human				x
AT3	SXB_BS.0056	15/01/2014	human				x
AU59	SXB_BS.0056	05/03/2014	human				x
BU81	SXB_BS.0204	05/02/2016	human		x	x	x
ER1089	SXB_BS.0204	14/05/2015	swine-muscle	x	x	x	x
ER1160	SXB_BS.0204	27/05/2015	swine- meat	x	x	x	x
ER1165	SXB_BS.0204	18/06/2015	swine- meat		x	x	x
ER1169	SXB_BS.0204	18/06/2015	swine- meat	x	x	x	x
ER278	SXB_BS.0204	20/11/2013	swine- meat	x	x	x	x
ER472	SXB_BS.0204	26/03/2014	swine-env*	x	x	x	x
ER822	SXB_BS.0204	25/11/2014	swine-carcass		x	x	x
N11	SXB_BS.0204	15/02/2012	swine-env*		x	x	x
U60	SXB_BS.0204	21/09/2012	swine-faeces-l.n.**		x	x	x
ER1860	SXB_BS.0204	01/12/2014	swine-carcass		x	x	
ER1861	SXB_BS.0204	18/09/2014	swine-carcass		x	x	
ER1865	SXB_BS.0204	01/08/2012	swine-salami		x	x	
ER1866	SXB_BS.0204	13/04/2015	swine-carcass		x	x	
ER1867	SXB_BS.0204	04/05/2015	swine-carcass		x	x	
ER1868	SXB_BS.0204	24/06/2015	swine-carcass		x	x	
ER792	SXB_BS.0204	28/10/2014	swine-sausages		x	x	
ER1858	SXB_BS.0204	01/03/2016	swine-carcass		x	x	
ER1859	SXB_BS.0204	09/02/2016	swine-carcass		x	x	
ER1869	SXB_BS.0204	29/10/2015	swine-meat		x	x	

AD45	SXB_PR.0071	01/02/2013	human				x
AT39	SXB_PR.0071	21/01/2014	human				x
ER386	SXB_PR.0071	11/02/2014	swine-carcass				x
H58	SXB_PR.0071	24/10/2011	human				x
I18	SXB_PR.0071	24/10/2011	human				x
Z57	SXB_PR.0071	19/10/2012	human				x
ER602	SXB_PR.0119	18/07/2014	no data		x	x	x
ER521	SXB_PR.0119	21/05/2014	swine		x	x	
U62	SXB_PR.0250	21/09/2012	swine-faeces-l.n.**		x	x	x
AR65	SXB_PR.0273	27/11/2013	human				x
AS3	SXB_PR.0273	03/12/2013	human				x
ER230	SXB_PR.0273	30/10/2013	swine-salami				x
ER312	SXB_PR.0273	06/12/2013	swine-salami	x			x
ER433	SXB_PR.0273	06/03/2014	swine-l.n.**				x
AH74	SXB_PR.0621	17/06/2013	swine-l.n.**		x	x	x
ER763	SXB_PR.0621	09/10/2014	swine		x	x	
ER1695	SXB_PR.0787	26/01/2016	swine-carcass		x	x	x
ER320	SXB_PR.0787	11/11/2013	swine-l.n**		x	x	x
BE8	SXB_PR.0932	28/10/2014	human				x
BF19	SXB_PR.0932	26/11/2014	human				x
ER542	SXB_PR.0932	05/06/2014	swine-carcass	x			x
ER819	SXB_PR.0932	25/11/2014	swine-carcass				x
ER841	SXB_PR.0932	22/12/2014	swine-carcass				x
ER737	SXB_PR.1076	23/09/2014	swine-carcass				x
ER1175	SXB_PR.1259	05/06/2015	swine-muscle		x	x	x
ER1289	SXB_PR.1274	12/06/2015	swine-sausages		x	x	x
ER1864	SXB_PR.1274	12/02/2014	swine-env*		x	x	

*env: environment-lairage; **l.n.: lynph-nodes.

Table 3: mutants generated in this work. For each strain, name and relevant features were reported.

Strain	Relevant features
ER1175ΔB1*	ΔB1::*cat*; CmR
ER1175ΔB2*	ΔB2::*cat*; CmR
ER1175ΔB3*	ΔB3::*cat*; CmR
ER1175ΔB4*	ΔB4::*cat*; CmR
ER1175ΔB5*	ΔB5::*cat*; CmR
ER1175ΔB6*	ΔB6::*cat*; CmR
ER1175ΔSGI1	ΔSGI1::*cat*; CmR
ER1175_PC	ER1175 plasmid cured
N11_PC	N11 IncFII-plasmid cured
ER1175::*hilD*_nv	Δ*hilD*_v::*hilD*_nv-*cat*; CmR
ER1175Δ*hilD*	Δ*hilD_v*::*cat*; CmR
ER1175Δ*invA*	Δ*invA*::*cat*; CmR
ER278::*hilD*_v	Δ*hilD*_nv::*hilD*_v-*cat*; CmR
ER278Δ*hilD*	Δ*hilD_nv*::*cat*; CmR
ER278Δ*invA*	Δ*invA*::*cat*; CmR

* B1-B6 and SGI-1: regions containing exclusive genes of isolates defined virulent based on their phenotype in *in vitro* cell culture infection assays.

2.2. *C. elegans* infection assays

Ten isolates were tested for virulence in *C. elegans* infection assay (Table 1): 5 isolates belong to pulsotype SXB_BS.0204, more prevalent in swine than in human, and 1 isolate to pulsotype SXB_BS.0056, as prevalent in swine as in human. The other 4 isolates belong to other PFGE profiles equally isolated from the two hosts. *Escherichia coli* OP50 was used as the nematode food source. *S.* Typhimurium LT2 was used as positive control of infection. Two *C. elegans* strains were used for the infection assays, SS104 (*glp-4*) and AU37 (*glp-4*; *sek-1*). SS104 is a temperature-sensitive mutant that produces progeny at 15°C but is sterile at 25°C. This feature allows for a constant

number of nematodes during the experiment and also avoids mixing adults with their progeny. AU37 is a temperature-sensitive mutant also defective for *sek-1* gene of the p38 mitogen-activated protein kinase (MAPK) pathway. It is one of the pathways regulating immunity of *C. elegans* and it has been shown to be involved in response to infection. AU37 strain is more sensitive to pathogens, but its survival is not significantly different from a wild type strain in a pathogen-free medium.

Worms were maintained at 15°C on NG agar medium (NGM) and were handled using standard techniques (79). For *C. elegans* infection assays each bacterial isolate was grown overnight, in LB medium, at 37°C with shaking (200 rpm). 50 µL of each bacterial culture were spotted on NGM in 60 mm diameter plates, and bacteria were allowed to grow for 24 h at room temperature. ~20 worms synchronized at L4 stage were placed on each plate and incubated at 25°C. A total of 3 plates (about 60 worms) were used for each *S*. Derby isolate tested.

Worm mortality was scored daily for 7 days, and a worm was considered dead when it failed to respond to a gentle touch with a wire-pick. Worms that died of protruding/bursting vulva, bagging and getting stuck to the wall of the plate were excluded from the analysis. Survival curves were obtained using the Kaplan-Meier method (80). Curve were compared by Log-Rank test (81) using the mean and maximum lifespan values. The Bonferroni correction was applied for multiple comparisons: two curves were considered significantly different when *p*-value (*p*) was $< 0.05/m$, where *m* is the number of total comparisons between curves made.

2.3. *In vitro* cell culture infection assays

A total of 39 *S.* Derby isolates were included in *in vitro* cell culture infection assays (Table 1): 20 isolates belong to SXB_BS.0204 profile and 9 isolates belong to SXB_BS.0056 profile. The other 10 isolates have 5 different PFGE pulsotypes, all closely related to SXB_BS.0204 and SXB_BS.0056:

similarity among these profiles is more than 80%. *S.* Typhimurium LT2 and *E. coli* BW25113 strains were used as positive and negative control of infection, respectively. Porcine IPEC-J2 and human INT-407 intestinal epithelial cell lines were used for the assays. IPEC-J2 cells were cultured in 50% Dulbecco's Modified Eagle's Medium (DMEM) and 50% Ham's F12 Nutrient Mixture containing 5% fetal bovine serum, penicillin 100 U/ml and streptomycin 100 µg/mL. INT-407 cells were propagated in DMEM containing 10% fetal bovine serum, penicillin 100 U/mL and streptomycin 100 µg/mL. Both cell lines were maintained at 37°C in a 5% CO_2 incubator. For the *in vitro* virulence assays IPEC-J2 and INT-407 cell lines were seeded in 96-well plates at $1.5x10^5$ and $4x10^5$ cell/mL respectively. Experiments were performed 48 hours post seeding to reach 100% confluency. Monolayers were washed twice with the specific medium of each cell line and they were incubated with antibiotics-free medium for 1 hour. For each assay, two monolayers were trypsinized to determine the mean number of epithelial cells per well, in order to calculate the amount of Salmonellae needed to reach a multiplicity of infection (MOI) of 10. Bacterial isolates were grown statically to stationary phase, at 37°C, in LB broth. *E. coli* strain BW25113 and S. Typhimurium LT2 were used as negative and positive control of infection, respectively. After inoculation, the multi-well plates were incubated at 37°C with 5% CO^2 for 1 h. After infection monolayers were washed twice and incubated for 1 h with fresh medium containing 100 µg/mL gentamicin to kill extracellular bacteria. To score the invasion efficiency of tested isolates, after the gentamicin treatment cell monolayers were washed two times and then disrupted with 0.1% Tryton X-100 treatment for 30 minutes. For the replication assay, medium containing 100 µg/mL gentamicin was replaced after 1 h with fresh medium containing 10 µg/mL gentamycin to maintain clearance of extracellular *Salmonella* in the medium. Plates were incubated for 20h and then monolayers were lysed. For each assay, after Tryton X-100 treatment, serial 10-fold dilutions of disrupted cell suspension were plated on LB agar and incubated at 37°C overnight for bacterial cell

count. Results were expressed as mean number of bacteria recovered per well divided by the inoculum. Each assay was performed in duplicate including two technical replicates for each biological replicate. Statistical analysis were made on log-transformed data to guarantee errors normality. The Monte Carlo Effects model (EM) was used to apply a linear mixed-effects model but allowing for censored normal responses when bacterial counts are below limit of detection (82). For each assay, limit of detection was calculated dividing the minimum number of countable bacteria per well (20 CFU) for the amount of bacterial cells used for the inoculum. All mutants generated were subjected to *in vitro* cell culture infection assays following the same protocol. Results were expressed as relative invasion and replication, i.e. the bacterial concentration relative to invasion and replication of each mutant normalized to the respective wild type isolate. Each assay was performed in triplicate including two technical replicates for each biological replicate. Statistical analysis was performed using as limit of detection the minimum number of countable bacteria per mL divided for mean concentration relative to invasion and replication of the wildtype. The Bonferroni correction was applied for multiple comparisons.

2.4. Screening for antibiotic resistance

S. Derby strains tested in virulence assays were screened for antibiotic resistance (Table 1). Each isolate was analyzed for resistance to antibiotics listed in Table 4. For each antibiotic, 2 different concentrations were tested: the epidemiological cutoff value indicated by the European Committee on Antimicrobial Susceptibility Testing (EUCAST, 83), when available, and the common working concentration for mutant selection. Only one concentration of Erythromycin was tested, because no epidemiological cutoff value was reported by EUCAST. *S*. Derby isolates were grown in cation-adjusted Muller-Hinton Broth (CAMHB) in 96-well plates, overnight at 37°C, statically. Each isolate was diluted in medium with antibiotic to a final concentration of 10^5 CFU/mL. After 18 h of

incubation at 37°C, the optical density at 600 nm (OD_{600}) was measured with a Synergy H1 Hybrid Multi-Mode Microplate Reader (BioTek, Winooski, VT, USA). To discriminate between resistant and sensitive isolates, the OD_{600} value of each isolate grown in presence of antibiotic was compared with the mean OD_{600} of *S.* Typhimurium LT2, calculated on 3 biological replicates. This strain is sensitive to every antibiotic used in this screening, therefore it was used as positive control when grown in CAMHB without antibiotics and as negative control when grown in presence of antibiotics. Z' value for each condition was calculated to assess the quality of assays (84). A strain was considered resistant to a specific antibiotic when its OD_{600} was above the lower confidence limit of the mean OD_{600} of *S.* Typhimurium LT2 grown without the antibiotic. The confidence interval was calculated as previously described (85), adding and subtracting to from the S. Typhimurium mean OD_{600} the standard error multiplied for 5.

Table 4: Antibiotics used for resistance screening. For each antibiotic but Erythromycin, 2 different concentrations were tested (expressed in µg/mL): the epidemiological cutoff value indicated by EUCAST and the common working concentration for mutants selection.

	EUCAST (µg/mL)	Mutants selection (µg/mL)
Ampicillin (AMP)	8	100
Gentamycin (GEN)	2	15
Tetracycline (TET)	8	20
Kanamycin (KAN)	2	50
Chloramphenicol (CM)	16	20
Spectinomycin (SPC)	2	30
Erythromycin (ERY)	-	20

2.5. Sequencing and bioinformatics

A total of 45 *S.* Derby isolates were Whole Genome Sequenced (WGS): 22 isolated had been tested by invasion and replication assays, whereas 23 isolates with PFGE profiles not closely related to those tested were added to the analysis to determine the relationship of the isolates of interest in

a broader context of *S*. Derby. Isolates were inoculated into Brain Heart Infusion (BHI) Broth and cultured overnight at 37°C with agitation (200 rpm). Genomic DNA was extracted by Qiagen DNeasy blood and tissue kit (Qiagen), following the kit instructions but extending the incubation with Proteinase K to 20 h to enhance cells lysis. Extracted DNA was quality controlled and quantified with a Take3 Micro-Volume plate, used in a Synergy H1 microplate reader. Sequencing libraries were prepared with the Nextera XT V2 sample preparation kit (Illumina, San Diego, CA, USA) and sequenced using the MiSeq Reagent Kit v2 on an Illumina MiSeq platform, with a 2x250-bp paired-end run. Each reads set was evaluated for sequence quality and read-pair length using FastQC software (86).

In order to find exclusive genes of one isolate or a group of them, selected based on the virulence phenotype, reads were *de novo* assembled using SPAdes version 3.12.0 (87) and protein-coding genes were predicted by Prodigal software tool version 2.6.3 (88). The OrthoMCL tool (89) was used to find classes of orthologous genes. OrthoMCL uses Markov clustering (MCL) on the results from an all-vs-all BLASTp to infer homologous (i.e., both orthologous and paralogous) relationships among a set of protein sequences. The output matrix, showing which genes are present or absent in each isolate, was analyzed by a custom script to select exclusive genes of a selected isolate or group of them. The genome of the selected isolate or one genome of each analyzed group was annotated by RAST version 2.0 (90). The identified exclusive genes were then mapped to the annotate genome representing the isolate or group analyzed, using Geneoius software, version 9.0.2 (https://www.geneious.com). This operation allowed genes annotation and localization on the reference genome.

PlasmidFinder version 2.0 (91) was used for the identification of plasmids in the sequenced isolates. Phaster (92) was used for the identification and annotation of prophage sequences. The sequence type (ST) of each isolate was determine directly from reads using the MLST service

available within EnteroBase (https://enterobase.warwick.ac.uk/species/senterica/allele_st_search).

Single Nucleotide Polymorphisms (SNPs) of each isolate were identified using SNIPPY (93): it is a pipeline that includes BWA MEM version 0.7.12 (94) to map reads against a reference genome, SAMtools version 1.1 (95) to convert SAM files to sorted and indexed BAM alignment files and Freebayes-0.9.20 (96) a bayesian genetic variant detector. U60 genome was used as reference: U60 was *de novo* assembled by SPAdes 3.11 and contigs with length > 1000bp were sorted by MAUVE (97) based on *S*. Typhimurium LT2 genome, chosen as a reference genome because no *S*. Derby complete genome was available at the time. They are then concatenated in a pseudochromosome which was annotated by RAST version 2.0 (90). A concatenated alignment composed of SNP sites from each sequenced isolate was generated and used as input for a maximum-likelihood phylogenetic tree generated by RAxML version 7.0.4 (98), starting from the core SNPs matrix, using a General Time-Reversible (GTR) substitution model with gamma correction for among-site rate variation. Support for nodes on the trees was assessed using 100 bootstrap replicates.

2.6. Construction of mutant strains

Gene deletion and allelic exchange were made using the Red recombinase system, based on methods described previously by Datsenko and Wanner (78). Plasmids used for construction of mutant strains are listed in Table 5. The recipient strains (*S*. Derby ER1175, ER278, N11 and *S*. Typhimurium LT2) were transformed with pKD46 plasmid carrying the Red recombinase genes.

Phusion High-Fidelity DNA Polymerase (Thermo Scientific) was used to amplify recombinant constructs. Thermo Scientific Tm Calculator was used to estimate the appropriate annealing temperature (T annealing) for each amplification reaction.

For gene deletion, entire coding sequences (CDSs) were replaced by Chloramphenicol resistance cassette (*cat*) amplified from pKD3 template plasmid or by Gentamycin resistance cassette (*gen*) amplified from pRU1098 plasmid (gift from Philip Poole - Addgene plasmid # 14463, 99). Primers to amplify the recombining amplicon were designed with 50-nucleotides extension homologous to regions immediately up- and downstream of the CDS, and 20 nucleotides homologous to the *cat*-resistance cassette (Table 6). The PCR products were gel purified with the PCR clean UP PCR extraction NUCLEO SPIN kit (Macherey-Nagel GmbH & Co).

Table 5: List of plasmids used in this study. Relevant features and source of each vector are specified.

Plasmid	Relevant features	Source
pKD46	bla PBAD gam bet exo pSC101 oriTS	(Datsenko and Wanner, 2000)
pKD3	bla FRT cat FRT PS1 PS2 oriR6K	(Datsenko and Wanner, 2000)
pRU1098	gent gfpmut3.1 LAA pbbr oriV	(Karunakaran et al, 2005)

In allelic exchange experiments, linear fragments were constructed to contain the gene allelic variant and the *cat*-resistance cassette fused at the 3'. For *hilD*, the region including both the ORF and the 3'UTR sequence was used, because it was demonstrated that *hilD* 3'UTR is a cis-acting element directly involved in hilD mRNA turnover (100). To obtain these constructs, the ORF or region to insert was amplified with the forward primer homologous to the 50-nucleotides sequence upstream the gene start codon and the reverse primer composed by 10 nucleotides homologous to the 5'end of the *cat*-resistance cassette and 10 nucleotides homologous to the 3' end of the ORF or region. The *cat*-resistance cassette was amplified using the forward primer containing 10 nucleotides homologous to the 3' end of the ORF or region and 10 nucleotides homologous to the 5'end of the *cat*-resistance cassette and a reverse primer designed with 50 nucleotides homologous to regions immediately downstream the ORF or region to insert and 20

nucleotides homologous to 3' end of the *cat*-resistance cassette. The two PCR products have therefore 20bp overlapping region, designed to have a T annealing of ~60°C. Once gel purified, the PCR products were assembled by Overlap PCR with the following cycle conditions, divided in two steps:

1st step	98°C	3 minutes	
	60°C	2 minutes	
	72°C	5 minutes	
2nd step	98°C	30 seconds	
	98°C	7 seconds	
	Specific T annealing	20 seconds	X 35 cycles
	72°C	1 minutes	
	72°C	7 minutes	

The PCR mix was assembled without primers, that were added at the end of the 1st PCR step.

The obtained linear fragments were purified and transformed into the recipient strain carrying pKD46 by electroporation.

Competent cells were prepared starting from an overnight culture grown in 3 mL LB broth at 37°C with 200 rpm shaking. The culture was diluted 1:100 in LB broth and grown to OD_{600} = 0.6. One hour before reaching this OD_{600} value, L-arabinose (Sigma) was added to a final concentration of 10 mM to induce expression of recombinase genes. Once pelleted, cells were washed twice with ice-cold H_2O and one time with 10% Glycerol. Cells were then suspended in 40 µL of 10% glycerol and mixed with DNA in a 0.1 cm electroporation cuvette (BioRad, Inc.). Cells were electroporated at 1.8 kV. Immediately after electroporation, 1 mL of SOC medium was added to the cuvette. Cells were recovered in SOC medium for 1–3 hours at either 30°C (for temperature-sensitive plasmids) or 37°C, with mild shaking (100 rpm). Cells were then concentrated ten-fold from a 1 ml culture and plated on selective media.

Transformants were selected on LB agar containing Cm 20 µg/mL or Gen 15 20 µg/mL. Insertion of linear fragments was verified by PCR using two pairs of primers specific for the *cat*-resistance

cassette and regions upstream and downstream of the target gene (Table 6). To verify the allelic exchanges, allelic-specific (AS) PCRs were set up to detect the allelic variant present in colonies positive to the previous PCR. Primers for AS PCR were designed by BatchPrimer3 (101) or WASP (102) (Table 6). A common reverse (or forward) primer and two forward (or reverse) AS primers were used for each SNP. The AS primers were designed with the 3' end complementary to each allele of a SNP. AS extension relies on the difference in extension efficiency of DNA polymerase between primers with matched and mismatched 3' ends. Two PCR reactions are needed to detect both alleles of a SNP. To guarantee specificity of allelic-specific primers, BatchPrimer3 and WASP incorporate a primer specificity enhancement: they introduce one 'mismatch' at the second- or third-last (at the 3'-end) base of AS primers to improve the resulting AS PCRs.

Table 6: primers used in this work, 5' → 3' sequences and purpose.

Primer	sequence 5' -> 3'	Purpose
CAT_Cswap_1_RV	CCGTTTTCACCATGGGCAAA	Control for *cat*-
CAT_Cswap_2_FW	CACGCCACATCTTGCGAATA	cassette insertion
		- *cat*-specific primers
GEN_Cswap_1_RV	AAGTTGGGCATACGGGAAGA	Control for *gen*
GEN_Cswap_2_FW	GGCTGATGTTGGGAGTAGGT	cassette insertion
		- *gen*-specific primers
B1_swap_FW	ATCTTGAGTATCTAATTACTCTCTGAACCATGAAAAATACCACGTTCA	B1 knockout
	AAGTGTAGGCTGGAGCTGCTTC	
B1_swap_RV	GTATCGAATTTATTTAACCCATTGTCCTGAAAGAAATTTTTTGAGGAT	
	ATGCCTACCTGTGACGGAAGAT	
B1_Cswap_1_FW	TGAATCGCCACGGATAACCT	Control for *cat*
B1_Cswap_2_RV	GCCTGCCGAAAGAGATGAAG	cassette insertion
		- B1-specific primers
B2_swap_FW	CACAGCAATGAAACTATTGCCTGTCTTTTCACCACTTCAGGCTCGGTG	B2 knockout
	GTGTGTAGGCTGGAGCTGCTTC	
B2_swap_RV	TTTCTACTTTCCAGTGATGCTTACCCACGGCTATCTCCTTAAAACTGT	
	GGGCCTACCTGTGACGGAAGAT	
B2_Cswap_1_FW	TGAAAGATGACCGGCAAAGC	Control for *cat*

B2_Cswap_2_RV	CCGGTTGAAATCACCCACAG	cassette insertion - B2-specific primers
B3_swap_FW	CTTATTGCCTTTCGTTCCGGCAAAAACATGAAGCATCTCGGAGGTGC TGCGTGTAGGCTGGAGCTGCTTC	B3 knockout
B3_swap_RV	TGTCAGCCCCGGATTAAGCTCGATAAACGCCTGCACTTTTGCTAACA GACGCCTACCTGTGACGGAAGAT	
B3_Cswap_1_FW	GTCTCTTGAATGCTCTGCCG	Control for *cat*
B3_Cswap_2_RV	GCAATCACCATCGGATCACC	cassette insertion - B3-specific primers
B4_swap_FW	AGACCCGTAGAGCAGGTTTATCTGGCAGGGCGATTTGTGGGGAACG TGGAGTGTAGGCTGGAGCTGCTTC	B4 knockout
B4_swap_RV	GTAGGTGTTAACGCTGGGTGGAAATGGCAATGTCAGTTTCATGCCG CCCCGCCTACCTGTGACGGAAGAT	
B4_Cswap_1_FW	GGTGATCCGATGGTGATTGC	Control for *cat*
B4_Cswap_2_RV	TTGAGATGGAAAGACGGGCT	cassette insertion - B4-specific primers
B5_swap_FW	TTAATTATTGATAGATTAGATGGGTATCGTTGAGGTGAAAATCAAAA GCAGTGTAGGCTGGAGCTGCTTC	B5 knockout
B5_swap_RV	ATATTAAAAAATGCGTTCTGCGTCAGTGTTCAAGGTATTGAGGAACC AATGCCTACCTGTGACGGAAGAT	
B5_Cswap_1_FW	CAGCCATTACGCCCAATCAA	Control for *cat*
B5_Cswap_2_RV	GCTATCTTCCCGGCAAACAG	cassette insertion - B5-specific primers
B6_swap_FW	GGAATCAGAATTAGGTGTACTATGTACTCCATTAAAAACAGGAGAC AGCCGTGTAGGCTGGAGCTGCTTC	B6 knockout
B6_swap_RV	TAAATTACACGCGAACGGCCTCGCATGTCGAATCCTTCTGTCACAAC GGCGCCTACCTGTGACGGAAGAT	
B6_Cswap_1_FW	GGTTTCAGGGCACGATCATC	Control for *cat*
B6_Cswap_2_RV	CGAATCCCAGCCAGAACATG	cassette insertion - B6-specific primers
SGI1_swap_FW	CCGACGACCTGCTGGGACGGATTTTCTCCAGCTTCTGTATTGGGAAG TGAGTGTAGGCTGGAGCTGCTTC	SGI-1 knockout
SGI1_swap_RV	TTACGGGTTTTGTAGGCCCGGTAAGCATTGCGCCACCGGGCAATAC AACGGCCTACCTGTGACGGAAGAT	
SGI1_Cswap_1_FW	GATCGCGTGCTGTTTATGGT	Control for *cat*
SGI1_Cswap_2_RV	GCTATCTTCCCGGCAAACAG	cassette insertion - SGI-1-specific primers
hok_swap_FW	AGATAGCCCCGTAGTAAGTTAATTTTCATTAACCACCACGAGGCATC	*hok* knockout

	CCTGTGTAGGCTGGAGCTGCTTC	
hok_swap_RV	CCAGCATCAGTCCTGAAAAGGCGGGCCTGCGCCCGCCTCCAGGTTG CTACGCCTACCTGTGACGGAAGAT	
hok_Cswap_1_FW	TGACCTGGAAACCCTGTTCT	Control for *cat* cassette insertion - *hok*-specific primers
hok_Cswap_2_RV	CCCGTACAACATCAGCAAGG	
yac_swap_FW	TCCTCTGCTGCAGCATGGCGGGATGAAATGAGCAGGAAGGTCGCCG GTAAGAAAAGCTGCTGACGGAACA	*yac* knockout
yac_swap_RV	CAGCTTATTGTTATTGATTATTATCATGCTGTGATATCCGGACAGCCC CTAAACGGATGAAGGCACGAAC	
yac_Cswap_1_FW	CGACGCTGAACTGAAGGATG	Control for *gen* cassette insertion - *yac*-specific primers
yac_Cswap_2_RV	AAACCACATTGCACATCCCC	
traG_FW	GGCAGGATTTTCCAGTGT	Control for N11-plasmid presence
traG_RV	GGGCACTTTCTGCTCTGT	
ColM_FW	TGGCGCTTGCTCACTATTTA	
ColM_RV	GAGCGACTCTCCGATAATGC	
repA_FW	GACCGACATCACGTTCACAC	
repA_RV	TTTACCGCCTCACCATTAGC	
MerT_FW	TGTCTGAACCACAAAACGGG	Control for INF-plasmid presence
MerT_RV	AACGGTCGATAGGGTTCCAG	
Tet_FW	TGACGGGCTGTTTCCTTTTG	
Tet_RV	GATCCTCGCCGAAAATGACC	
Tn21_FW	CGTTCCGTCTTGCCATCAAT	
Tn21_RV	TTTCAGATGTTGCAGGCAGG	
Str_FW	AGACATCATGAGGGTAGCGG	Control for streptomycin resistance presence
Str_RV	CGGCCACAGTAACCAACAAA	
ydiV_1_FW	GTCCGTAGTGACTGGATGGCGAATAGCGCCCTAACCATGGGACTGG CGTA	*ydiV* Overlap-PCR
ydiV_2_RV	CAGGTAGGCTTATCGCTGAACGAGTTTAA	
ydiV_3_FW	TCAGCGATAAGCCTACCTGTGACGGAAGAT	
ydiV_4_RV	TCTTCAGTGTAGACGGTTAATCACCGGTTAAACACCGGCAAACAGAA AGGGTGTAGGCTGGAGCTGCTTC	
ydiV_Cswap_1_FW	GAATTGCTGCCCGGAGATTT	Control for *cat* cassette insertion - *ydiV*-specific primers
ydiV_Cswap_2_RV	AACGTCCCTGATGATCGGAA	
ydiV_SNP_FW	GATGAAAACGCTCGGCTTGT	*ydiV* allelic specific

ydiV_SNP_noninf_RV	ACGTTTCATTGTTTTTGCCATA	PCR
yhaK_1_FW	GCAGCGTCCAGCCCCTTAACATAAAAGGAAGTAAAGAGAGGTTAATAACG	*yhaK* Overlap-PCR
yhaK_2_RV	CAGGTAGGCCTACACCGGTAAATCTACCA	
yhaK_3_FW	CCGGTGTAGGCCTACCTGTGACGGAAGAT	
yhaK_4_RV	ACGTTTCTTCGCCGATTTTTTGCTCATCTTGCGACTCCTTTTCTGGCTGGGTGTAGGCTGGAGCTGCTTC	
yhaK_Cswap_1_FW	GTCGGTACCAGAGCTTCAGT	Control for *cat* cassette insertion - *yhaK*-specific primers
yhaK_Cswap_2_RV	CGTACTGATGGCGTTGGATG	
yhaK_SNP_noninf_FW	GTTAACCCGAATGCAACTGTCA	*yhaK* allelic specific PCR
yhaK_SNP_RV	GCCCGTGAAGCTGGAAATTA	
hilD_1_FW	CAGTAGGATACCAGTAAGGA	*hilD* Overlap-PCR
hilD_2_RV	CAGGTAGGCCGTGACTGTTTCGGTGTAGA	
hilD_3_FW	AACAGTCACGGCCTACCTGTGACGGAAGAT	
hilD_4_RV	ATCGATATCAGGAGGAAGAGAAGAGGTATGCCTGGCAGAAAGCTAACAAGGTGTAGGCTGGAGCTGCTTC	
hilD_Cswap_1_FW	ATCAGACCATTGCCAACACA	Control for *cat* cassette insertion - *hilD*-specific primers
hilD_Cswap_2_RV	GGGGGTGTAAATGCTGCTTA	
hilD_SNP_noninf_FW	ACAGCACGTCCTACTTCATTCAAGA	*hilD* allelic specific PCR
hilD_SNP_inf_FW	ACAGCACGTCCTACTTCATTCAAGG	
hilD_SNP_RV	CGTGACTGTTTCGGTGTAGA	
hilD_KO_FW	CAGTAGGATACCAGTAAGGAACATTAAAATAACATCAACAAAGGGATAATGCCTACCTGTGACGGAAGAT	*hilD* knockout
invA_KO_FW	GTCGTACTATTGAAAAGCTGTCTTAATTTAATATTAACAGGATACCTATAGCCTACCTGTGACGGAAGAT	*invA* knockout
invA_KO_RV	TAATTCAGCGATATCCAAATGTTGCATAGATCTTTTCCTTAATTAAGCCCGTGTAGGCTGGAGCTGCTTC	
invA_Cswap_1_FW	ACAGGCGGAAATACAACAGG	Control for *cat* cassette insertion - *invA*-specific primers
invA_Cswap_2_RV	GAAATCGGAAGTGGCAAAAA	
gmk_RTPCR_FW	TTGGCAGGGAGGCGTTT	Real Time PCR on RNA extracted during epithelial cell infection
gmk_RTPCR_RV	GCGCGAAGTGCCGTAGTAAT	
hilD_RTPCR_FW	GGCGCTCTCTATGCACTTATC	
hilD_RTPCR_RV	GCAGGAAAGTCAGGCGTATAG	

hilA_RTPCR_FW	ATCGTCGGGAGTTTGCTATTC	
hilA_RTPCR_RV	CTGACCAGCCATGAAGAGATT	
hilC_RTPCR_FW	TGTCCACGGGTTTGTAGTAATG	
hilC_RTPCR_RV	TGCTCGCTCAAGGAAATCAA	
rtsA_RTPCR_FW	CAGGTGGGGAGCATTGAATG	
rtsA_RTPCR_RV	GGTGAGCTTGATGAGTACGG	
invA_RTPCR_FW	TGGAGCATATTCGTGGAGCA	
invA_RTPCR_RV	AGGTCTGACGGATCCCTTTG	
invF_RTPCR_FW	AGAAGGCCACGAGAACATCA	
invF_RTPCR_RV	GAAACGCCGATCAGCTCTTT	
sipB_RTPCR_FW	GCAAAATGATGGGCGAAACG	
sipB_RTPCR_RV	CATTACCCAGGCCGCTAGTA	
pagN_RTPCR_FW	CTGCATCATTCCCCTTGTGG	
pagN_RTPCR_RV	AAAACGCCTTTGGTACGGTC	
Gapdh_RTPCR_FW	TGTAGACCATGTAGTTGAGGTCA	Real Time PCR on
Gapdh_RTPCR_FW	AGGTCGGTGTGAACGGATTTG	RNA extracted from
cxcl1_RTPCR_FW	ATGGCTGGGATTCACCTCAA	mice cecal tissue
cxcl1_RTPCR_RV	AGTGTGGCTATGACTTCGGTTT	
nos2_RTPCR_FW	TTGGGTCTTGTTCACTCCACGG	
nos2_RTPCR_RV	CCTCTTTCAGGTCACTTTGGTAGG	

2.7. Plasmid curing

In order to eliminate plasmids carried by N11 and ER1175 (i.e. virulent strain), these strains were treated with acridine orange. As N11 plasmid carries two toxin/antitoxin systems that induce host killing upon plasmid loss, toxin genes were knocked out by recombination before plasmid curing. Isolates were grown overnight in LB broth at 37°C, with 220 rpm shaking, then diluted to 10^3 CFU/mL and treated with acridine orange 150 µg/mL. After 48 hours of incubation, cultures were 10-fold serially diluted and plated on LB agar. As plasmids to be cured carry antibiotic resistance cassettes, plasmid-cured colonies were negatively selected on LB plates containing the relative antibiotics. Plasmid curing was then verified by PCR using primers designed on plasmid exclusive genes (Table 6).

2.8. RNA extraction at the early stationary phase, Preparation of cDNA Libraries and Illumina Sequencing

For RNA-seq experiment, RNA was extracted during early stationary phase because it was reported that in this phase SPI-1 genes and the other *hilD*-regulated genes are significantly up-regulated in *S*. Typhimurium wildtype compared to a mutant deleted for *hilD* (103). Strains were grown overnight in 5 mL LB broth and diluted to OD_{600} = 0.1 in 250 mL flasks in 20 mL LB broth. Flasks were incubated at 37°C and 220 rpm until OD_{600} = 2.0, which coincide to early stationary phase. 0.6 mL of each culture (about 1.2 x 10^9 cfu/mL) was centrifuged at maximum speed for 2 minutes and pellets were frozen in a dry ice ethanol bath. RNA extraction was performed by FastRNA SPIN Kit for Microbes (MP Biomedicals) according to the protocol. Briefly cells, resuspended in Lysis Buffer, were processed in a FastPrep-24 Instrument for 90 seconds, at 6.0 m/s, for cell disruption. SPIN columns were used for the purification process. Each sample was quantified and quality checked by 260/280 an 260/230 ratios using NanoDrop 2000 (Thermo Scientific). Trace DNA was removed by DNase digestion with TURBO DNA-free kit (Invitrogen Ambion), using rigorous treatment. RNA samples were then precipitated for purification: 0.1 volume 3M sodium acetate pH 5.2, 1 µL glycogen and 2.5 volumes 100% ice-cold ethanol were added to each sample that was then incubated at -20°C overnight. RNA pellets were washed two times in 75% ethanol, air-dried for 10 minutes at room temperature and resuspended in water. RNA quality was assessed using a 2100 Bioanalyzer (Agilent) according to Agilent RNA 6000 Nano Kit protocol and each sample was quantified by Qubit following the Qubit RNA BR (Broad-Range) Assay Kit protocol. Complete loss of genomic DNA after DNase treatment was verified by PCR on RNA samples using primers designed on RNA polymerase, sigma 70 factor sequence (Table 6). Each sample was then retro-transcribed by SuperScript VILO cDNA Synthesis Kit (Invitrogen), after normalization to 10 ng/µL directly into the cDNA synthesis master mix. In order to obtain double-

strand cDNA for libraries preparation, single-strand cDNA samples were firstly purified using Agencourt AMPure XP Beads (Beckman Coulter), then the second cDNA strand was synthesized using NEBNext Ultra II Non-Directional RNA Second Strand Synthesis. A second step of beads-based purification was performed before libraries preparation. 1.8X beads concentration was used in both purification steps to maximize recovery of smaller cDNA molecules (up to ~100bp length). Each sample was quantified by Qubit and normalized to the concentration required for starting library preparation (0.2 ng/µL). RNA-Seq libraries were prepared using Nextera XT V2 kit according to the reference guide. For the normalization step, each library was quantified by Qubit, analyzed by 2100 Bioanalyzer using a High Sensitivity DNA kit for quality control and to determine library size. Each library concentration was then adjusted to 2 nmol/L. 5µL of each library were pooled together and the obtained pool was quantified by Qubit. Denaturation and final dilution steps were made using the amount of reagents indicated for 2 nM Starting Library. NextSeq 500/550 Mid Output v2 kit (150 cycles) was used for the sequencing run. The current benchmarks recommended by the Encyclopedia of DNA Elements (ENCODE) Project are 70-fold exome coverage for standard RNA-Seq and up to 500-fold exome coverage to detect rare transcripts and isoforms (104). As the Mid-Output Flow cell with up to 260 M paired-end reads makes 16-19 Gb output, and 18 samples were sequenced, each one should have about 1 Gb output, assuming to obtain the same sequencing coverage for each sample. Salmonella exome is ~5MB, so ~1GB of data corresponds to ~200 X coverage. The instructions in NextSeq 550 system guide were followed to set up the sequencing run. Three independent biological replicates of RNA from each strain were sequenced.

2.9. Mapping of Sequenced Reads and Differential Gene Expression

For differential expression analysis, the FASTQ sequencing data files generated by the NextSeq for

each sample were merged and reads were adapter-trimmed and quality-checked by Trim Galore (105). To generate the reference for RNA-seq analysis, ER1175 genomic reads were assembled by SPADEs and contigs were concatenated and annotated by RAST. In order to find and remove genes absent in ER278, ER278 genomic reads were mapped to ER1175 concatenated contigs by minimap (106) and bedtools (107). ER1175 genes less than 99% covered by ER278 reads were removed from differential expression analysis. Transcriptomic Reads from each sample were mapped to ER1175 draft genome by HISAT (108). SAM files obtained were sorted and indexed by SAMtools to generate BAM alignment files. HTseq (109) was used to estimate the number of reads for each sample mapped on each annotated gene. Detection of differentially expressed genes was made using the package DESeq2 (110): it performs the normalization of reads count for each sample, the dispersion estimation (i.e. within group variability) and the log-fold change estimation. The package then applies single generalized linear model to detect genes significantly up- or down-regulated in different samples. DESeq2 includes multiple testing correction (Banjamini-Hochberg correction) and low-count gene filtering.

2.10. RNA extraction during epithelial cell infection and RT-qPCR

To extract total Salmonella RNA during epithelial cell infection, the same protocol described for invasion assay was used with some modifications. Treatment with gentamycin was removed to collect both adhered and intracellular bacteria. RNA was extracted after 0 (T0), 30 (T30) and 60 (T60) minutes of infection. For T0 each strain culture was diluted to get a MOI 10 and RNA was extracted without incubation with epithelial cells. 10^7 INT-407 cells were seeded on a 75 cm^2 tissue culture flask and incubated 48 hours at 37°C and under 5% CO^2. The overnight bacterial cultures were diluted in DMEM containing 10% fetal bovine serum and used to infect the epithelial cell monolayer using a MOI of 10. After 30 or 60 minutes of infection each flask was washed 5 times

with PBS and 5 ml of ice cold 'eukaryotic cells lysis and RNA stabilization solution' (0.1% SDS, 19% ethanol, 1% acidic phenol in water) was added. Flasks were incubated on ice for 30 minutes, then intracellular and adhered Salmonella were collected by centrifugation (27,500g for 10 minutes at 4°C). RNA was extracted from three biological replicates for each strain by RNeasy Mini Kit (Qiagen), after 1 hour of incubation of each pellet with lysozyme and proteinase K at a final concentration of 2 mg/ml and 15 mg/mL, respectively. During RNA purification, RNase-Free DNase Set (Qiagen) was used to add a step of on-column digestion of trace DNA potentially contaminating the sample. High-Capacity cDNA Reverse Transcription Kit (Thermo Scientific) was used to generate complementary DNA (cDNA) for Real-time reverse-transcription quantitative PCR (RT-qPCR). mRNA expression of 8 virulence genes was analyzed (primers listed in Table 6). All RT-qPCR reactions were performed using SYBR Green JumpStart *Taq* ReadyMix (Sigma). A total reaction volume of 15 µL was used, to include: 7.5 µL of 2X SYBR Green, 0.3 µL of (25 µmol/L) forward primer, 0.3 µL of (25 µmol/L) reverse primer, and template cDNA volume necessary to get 2 ng per reaction. Quantitative PCR was performed on a CFX96 Touch Real-Time PCR Detection System (bio-rad), using cycling and dissociation curve parameters according to the SYBR GREEN kit and instrument recommendations. Each RT-qPCR reaction was performed in three technical replicates for each sample. Expression of target genes were normalized to the reference house-keeping gene *gmk* and the relative transcriptional levels were analyzed through the $2^{-\Delta\Delta Ct}$ method correct for exact PCR efficiencies and the mean crossing point deviation between sample and control. ER1175wt - T0 was used as control condition. Subsequently differences in genes expression among strains were tested for significance by a Pair Wise Fixed Reallocation Randomisation Test. Calculations were made using the Relative Expression Software Tool (REST2009; Qiagen) (111, 112).

2.11. Mice experiments

Pathogen-free ~8weeks old C57BL/6 mice were used for infection experiments. They carry mutations in the Natural Resistance-Associated Macrophage Protein 1 (NRAMP) gene, developing a disease similar to typhoid fever when infected by S. Typhimurium (113). For the experiments, animals were housed in groups of five, under standard barrier conditions in individually ventilated cages. One group was used for each tested bacterial strain. Mice were pretreated by gavage with 20 mg of streptomycin resuspended in 200 µL of water to eliminate commensal intestinal bacteria (114). Twenty-four hours after streptomycin pretreatment, mice were intragastrically inoculated with 5 x 10^6 CFU of *S.* Derby dispersed in 2 ml PBS. Each inoculum was prepared by growing strains overnight in 50 ml LB medium, in 250 mL flask, at 37°C with 200 rpm shaking. Bacterial concentration in the inoculum was determined by spreading serial 10-fold dilutions on LB agar plates. 5 mice treated with 2 mL of PBS by gavage were used as negative control of infection. 3 days post-infection, mice were humanely sacrificed by CO_2 asphyxiation, and whole ceca, spleens, and livers were removed for mRNA purification, histopathology analysis and for bacterial count.

2.11.1 Analysis of bacterial loads in cecum, spleen, and liver

Because of the presence of microbiota in mice gut despite the streptomycin treatment, it was necessary to detect an antibiotic that selectively allows *S*. Derby strains growth and inhibits fecal microbiota. It was known, from antibiotic resistance screening, that all *S*. Derby strains used in the experiment are naturally resistant to Erythromycin at 20 µg/mL. In order to find the erythromycin minimum concentration needed to prevent feces microbiota growth without interfere with S. Derby growth, *Salmonella* strains and mice feces were plated on LB + the following antibiotic concentrations: 5,10, 20, 40, 80, 160, 200, 250 µg/mL. 200 µg/mL was found to be the best concentration for this purpose, therefore LB + 200 µg/mL Erythromycin plates were used to

determine bacterial loads in mice cecum, spleen and liver.

All tissue samples were weighted, manually mashed and homogenized in a Stomacher 80 micro-Biomaster adding 5 mL sterile 1 x PBS. Each homogenate was 10-fold serial diluted and plated on LB agar + 200 µg/mL Erythromycin. Plates were incubated overnight at 37°C. Differences between bacterial loads in each organ were evaluated Mann-Whitney U test with Bonferroni correction was applied for multiple comparisons.

2.11.2 Cecal RNA extraction and Quantitative Real-Time RT-PCR Analysis

Total cecal RNA was extracted for the analysis of expression of host genes correlated to gut inflammation. Mouse cecal sections were homogenized in 1 ml TRI Reagent (Molecular Research Center) by Bertin Precellys 24 Bead Tissue Homogenizer. After adding 0.2 mL of chloroform to the supernatant, phases were separated in 5PRIME Phase Lock Gel tubes. The aqueous phase was then isolated and RNA was purified. Trace DNA was removed by DNase digestion with TURBO DNA-free kit, using rigorous treatment. RNA samples were then precipitated as already described. RNA quality was assessed by 260/280 and 260/230 ratios using a NanoDrop 2000 (Thermo Scientific) and each sample was quantified by Qubit following the Qubit RNA BR (Broad-Range) Assay Kit protocol. Complete loss of genomic DNA after DNase treatment was verified by PCR on RNA samples using primers designed on RNA polymerase, sigma 70 factor sequence. Each sample was then reversely transcribed using SuperScript VILO cDNA Synthesis Kit (Invitrogen). Then, SYBR Green based real-time PCR was performed on target genes (*Cxcl1* and *Nos2*) using primers listed in Table 6. *Cxcl1* encodes for a chemoattractant of neutrophils, the first immune cells to migrate into infected tissue sites. *Nos2* encodes for a nitric oxide synthase induced during gut inflammation: the consequent production of nitric oxide (NO) represents an important microbicidal mechanism exploited by the innate immune system. The expression of these genes was monitored to detect presence of cecal inflammation, as previously performed (115).

All RT-qPCR reactions were performed using QuantiNova SYBR Green PCR Kit, according to the manufacturer instruction. A total reaction volume of 25 µL was used, to include: 7.5 µL of 2X QuantiNova SYBR Green PCR Master Mix + ROX Reference Dye, 1 µL of (14 µmol/L) forward primer, 1 µL of (14 µmol/L) reverse primer, and 1 µL cDNA (normalized to 5ng/µL). Quantitative PCR was performed on StepONE Plus (Applied Biosystems), with cycling and dissociation curve parameters used according to the SYBR GREEN kit and instrument recommendations. Each RT-qPCR reaction was performed in three technical replicates for each sample. mRNA expression levels of target genes were normalized to the reference gene Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and the relative quantification was obtained by $2^{-\Delta\Delta Ct}$ method correct for exact PCR efficiencies and the mean crossing point deviation between sample and control. Uninfected mice were used as control condition. Differences in genes expression were tested for significance by a Pair Wise Fixed Reallocation Randomization Test. Calculations were made using REST2009 software.

3. Results

3.1. Similar virulence between SXB_BS.0204 and SXB_BS.0056 isolates observed by using the *C. elegans* infection model

The epidemiological data on *S*. Derby, collected by the Emilia Romagna Enter-net reference center, revealed that the two most prevalent PFGE profiles in swine have different prevalence in humans: SXB_BS.0204 is significantly less frequently isolated from humans than from swine (SR significantly lower than 1), whereas SXB_BS.0056 is proportionally isolated from humans as well as from swine (SR not significantly different from 1). To evaluate whether the different distribution of these PFGE profiles in the two hosts correlates with a different virulence level, we performed infection assays in *C. elegans*. This nematode has already been used as infection model for many pathogens, including *Salmonella*. In particular, it was shown that several *S*. Typhimurium mutants deleted for genes know to be involved in human pathogenesis exhibit diminished ability to kill the worm (116, 117, 118).

Isolates included in the infection assays belong to SXB_BS.0204 and SXB_BS.0056 profiles, as well as to PFGE profiles other than SXB_BS.0056, but with SR similarly not different from 1 (Table 7).

Table 7: selection ratio (SR) and relative *p*-value for each pulsotype considered

Pulsotype	SR	*p*-value
SXB_BS.0204	0.0875	0.00009
SXB_BS.0056	1.458333	0.08983
SXB_PR.0932	2.25	0.05201
SXB_PR.0073	1.68	0.02831
SXB_PR.0273	0.525	0.11617
other	0.993932	0.49270

Two *C. elegans* strains (SS104, temperature sensitive, and AU37, sensitive to temperature and pathogens infection) were fed on lawns of the selected *S*. Derby isolates. *Escherichia coli* OP50 and *S.* Typhimurium LT2 were used as negative and positive control of infection, respectively. Fig. 10

and 11 show the survival curves relative to SS104 and AU37 strains infection assays, respectively. Table 8 and 9 include *p*-value for each comparison, obtained applying the Log-Rank Test and correcting the threshold value for multiple comparisons (= 0.000758). No significative differences in survival rates were found among tested *S*. Derby isolates and LT2 in SS104 infection. For what concerns AU37 infection results, all S. Derby isolates are associated to survival rates higher than that of LT2. Only sporadic differences between isolates were detected, but no correlation between survival rate and PFGE profile was detected.

Using this infection model, SXB_BS.0204 isolates did not show significant differences in virulence compared to the SXB_BS.0056 isolate and the other isolates with pulsotypes similarly found with the same prevalence in human and swine.

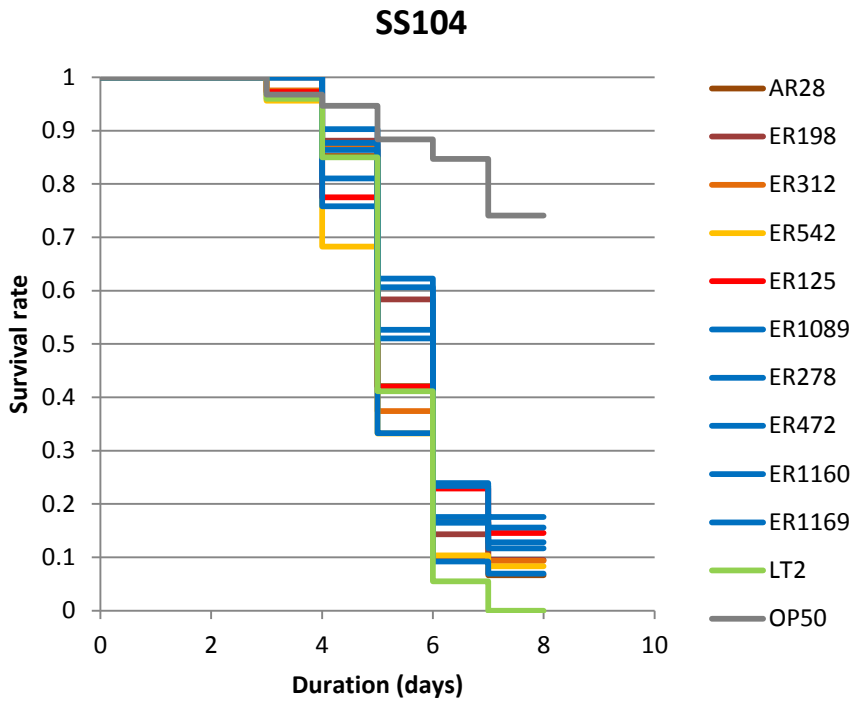


Figure 10: Survival curves of SS104 *C. elegans* strain infected with SXB_PR.0073 (brown), SXB_PR.0273 (orange), SXB_PR.0932 (yellow), SXB_BS.0056 (red) and SXB_BS.0204 (blue) isolates, S. Typhimurium LT2 (green) and E. coli OP50 (grey), monitored for 7 days.

Table 8: Log-Rank test derived *p*-values for multiple pairwise comparisons of survival curves of SS104 strain. Underlined values are below the threshold value, corrected for multiple comparisons (threshold=0.000758).

	AR28	ER198	ER312	ER542	ER125	ER1089	ER278	ER472	ER1160	ER1169	LT2	OP50
AR28	1	0.0678	0.7805	0.1230	0.6835	0.9054	0.1179	0.0206	0.0820	0.3532	0.1380	**1.53E-11**
ER198		1	0.0891	0.0075	0.4280	0.1144	0.8317	0.6338	0.9425	0.3742	0.0049	**1.28E-09**
ER312			1	0.1830	0.7139	0.3243	0.0640	0.1746	0.1016	0.2668	0.4374	**6.86E-11**
ER542				1	0.1371	0.6947	0.0070	0.0012	0.0943	0.0302	0.6445	**1.90E-11**
ER125					1	0.2072	0.3286	0.1188	0.2969	0.6147	0.1241	**8.86E-09**
ER1089						1	0.2025	0.0698	0.2068	0.4551	0.8404	**1.10E-11**
ER278							1	0.4966	0.8856	0.5275	0.0031	**2.77E-08**
ER472								1	0.6593	0.2066	0.0003	**3.31E-07**
ER1160									1	0.5117	0.0039	**1.09E-07**
ER1169										1	0.0323	**7.57E-09**
LT2											1	**1.13E-14**
OP50												1

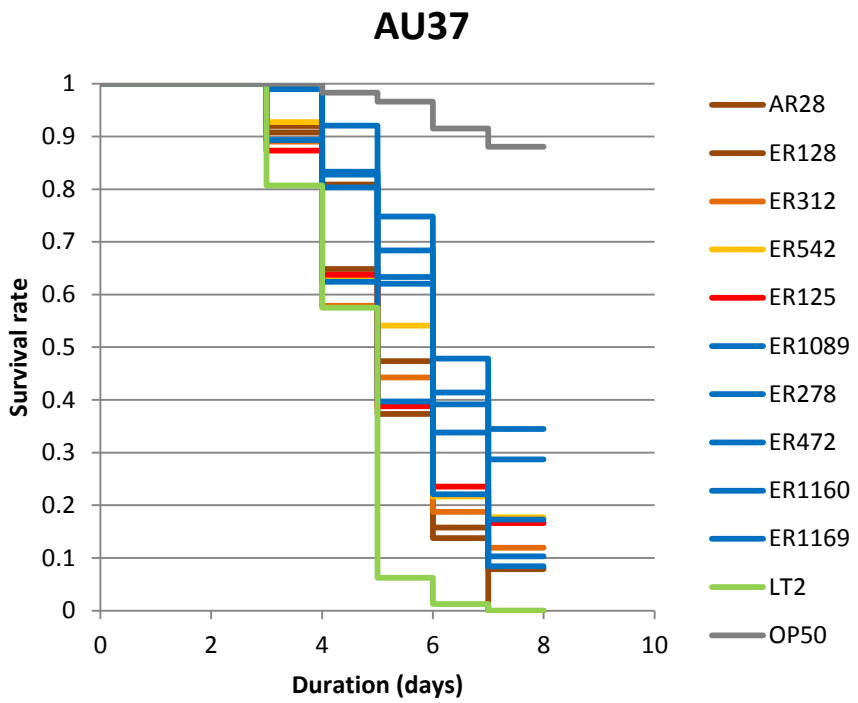


Figure 11: Survival curves of AU37 *C. elegans* strain infected with SXB_PR.0073 (brown), SXB_PR.0273 (orange), SXB_PR.0932 (yellow), SXB_BS.0056 (red) and SXB_BS.0204 (blue) isolates, S. Typhimurium LT2 (green) and E. coli OP50 (grey), monitored for 7 days.

Table 9: Log-Rank test derived *p*-values for multiple pairwise comparisons of survival curves relative to SS104 strain. Underlined values are below the threshold value, corrected for multiple comparisons (threshold=0.000758).

	AR28	ER198	ER312	ER542	ER125	ER1089	ER278	ER472	ER1160	ER1169	LT2	OP50
AR28	1	0.0839	0.4028	0.0431	0.2068	0.3440	**0.0001**	0.0337	**0.0002**	0.0050	**0.0006**	**4.90E-15**
ER198		1	0.4672	0.5865	0.7690	0.4946	**0.0005**	0.0017	0.0126	0.1656	**1.56E-07**	**4.78E-15**
ER312			1	0.2670	0.7006	0.9661	0.0003	**0.0006**	0.0078	0.1234	**0.0002**	**9.15E-15**
ER542				1	0.4414	0.2311	0.0218	0.0144	0.0723	0.5636	**4.14E-07**	**6.04E-12**
ER125					1	0.6287	0.0027	0.0032	0.0199	0.2361	**7.21E-05**	**1.48E-17**
ER1089						1	**0.0002**	**0.0003**	0.0048	0.1040	**9.69E-05**	**3.45E-20**
ER278							1	0.3889	0.4754	0.2010	**2.72E-20**	**5.93E-16**
ER472								1	0.9117	0.4597	**2.72E-13**	**1.89E-13**
ER1160									1	0.5029	**1.02E-08**	**2.25E-10**
ER1169										1	**2.04E-08**	**2.07E-14**
LT2											1	**5.09E-29**
OP50												1

3.2. SXB_BS.0204 and SXB_BS.0056 isolates differ in their ability to infect human and swine epithelial cell lines

The epidemiological differences observed between SXB_BS.0204 and SXB_BS.0056 isolates do not correlate with differences in virulence in *C. elegans* infection. Sem and Rhen, 2012 (119) showed that virulence of *S. enterica* in the nematode does not appear to involve the classical invasive and intracellular phenotype of the pathogen, but rather the ability to provoke overwhelming systemic oxidative stress in the host. Hypothesizing that differences in virulence between SXB_BS.0204 and SXB_BS.0056 isolates could reside in the ability to infect host cells, *in vitro* cell line infection assays were performed. Human (HU) Hela-derivative INT-407 and swine (SW) intestinal epithelial IPEC-J2 cell lines were used to perform invasion and replication assays. 39 Isolates belonging to SXB_BS.0204, SXB_BS.0056 and closely related profiles (CRPs, similarity>80%) are included in the assay (Table 2). Results were shown as invasion and replication efficiencies, i.e. mean number of bacteria recovered per well in invasion and replication assay divided by the inoculum (Fig. 3, a and b), and analyzed by linear mixed-effects model, allowing for censored normal responses when ratios are below limit of detection.

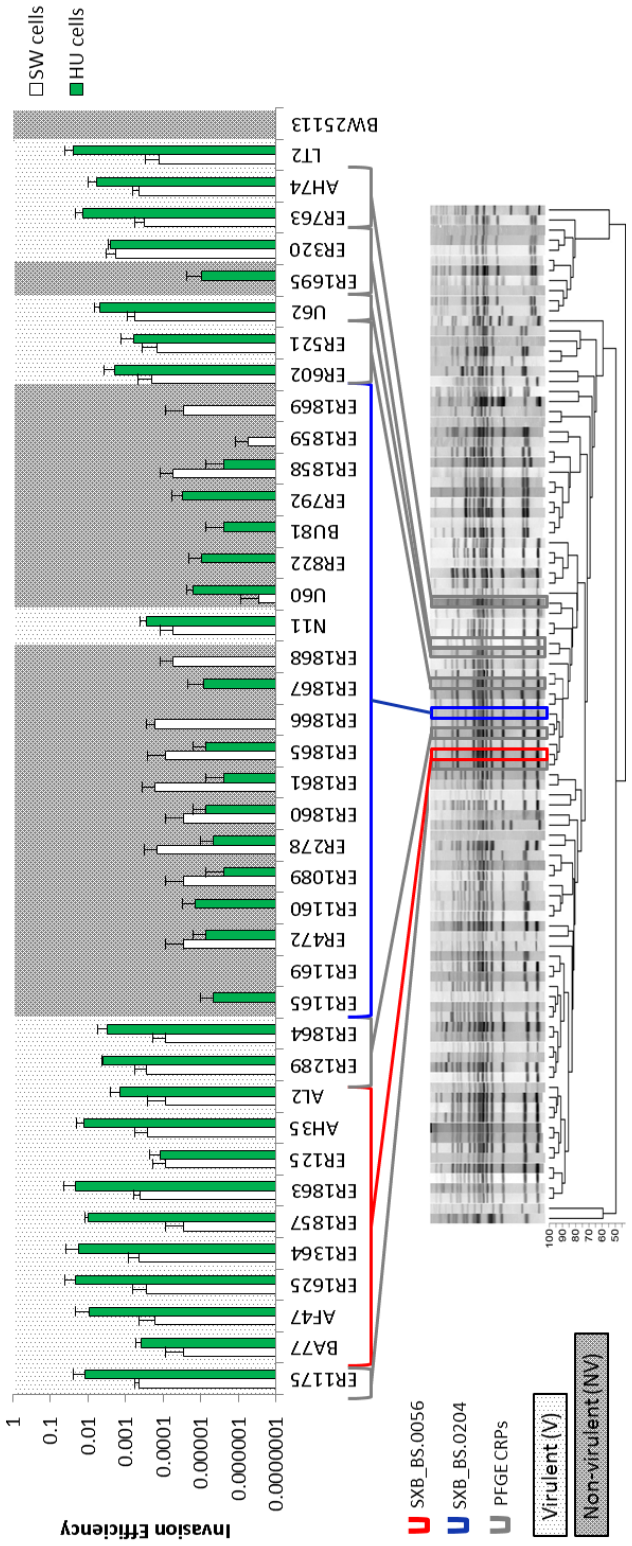
The highest differences in virulence levels were observed in HU cell infection: the mean invasion and replication efficiencies of SXB_BS.0204 isolates (invasion=0.0000094, replication=0.000022) are respectively 3 and 4 logs lower than those of isolates belonging to SXB_BS.0056 and CRPs (invasion=0.0047, replication=0.11), resulting in a significant difference (p<0.001). We next assessed if, among SXB_BS.0204 isolates, there were isolates with a different virulence phenotype. Invasion and replication efficiencies of N11 isolate (invasion=0.00025, replication=0.0010) result significantly higher than those of the other SXB_BS.0204 isolates (invasion=0.000010 , replication= 0.000022, p<0.05). Simultaneously, we tested if among SXB_BS.0056 and CRPs isolates there were isolates showing different virulence level: invasion and replication efficiencies of ER1695 isolate

(invasion=0.000011, replication=0.000020) was significantly lower than those of the other SXB_BS.0056 isolates (invasion=0.0048, replication=0.11, $p<0.05$). The same statistical analysis was performed on IPEC-J2 cell line infection results. The invasion and replication efficiencies of SXB_BS.0056 and CRPs isolates (invasion=0.00018, replication = 0.0051) are significantly higher than those of SXB_BS.0204 isolates (invasion=0.000046, replication=0.0010, $p<0.001$), even if the difference, less than 1 log, is lower than in human cells. Invasion and replication efficiencies of N11 isolate (invasion=0.000086, replication=0.0017) are not significantly different from that of the other SXB_BS.0204 isolates (invasion=0.000060, replication=0.00098, $p>0.05$). ER1695 infects swine cells at a lower extent (invasion=0.0000010, replication=0.00032) than SXB_BS.0056 and CRPs isolates (invasion=0.00018, replication=0.0055, $p<0.05$).

Isolates were divided in two groups based on their invasion and replication efficiencies in human cells:

- Virulent (V) isolates: this group includes isolates with SXB_BS.0056 pulsotype, N11 (SXB_BS.0204 pulsotype), and CRPs isolates except ER1695.
- Non-virulent (NV) isolates: this group includes isolates with SXB_BS.0204 pulsotype, except N11, and ER1965 from CRPs.

a


SW cells
HU cells
Invasion Efficiency
1
0.1
0.01
0.001
0.0001
0.00001
0.000001
0.0000001
BW25113
LT2
AH74
ER763
ER320
ER1695
U62
ER521
ER602
ER1869
ER1859
ER1858
ER792
BU81
ER822
U60
N11
ER1868
ER1867
ER1866
ER1865
ER1861
ER1860
ER278
ER1089
ER1160
ER472
ER1169
ER1165
ER1864
ER1289
AL2
AH35
ER125
ER1863
ER1857
ER1364
ER1625
AF47
BA77
ER1175
SXB_BS.0056
SXB_BS.0204
PFGE CRPs
Virulent (V)
Non-virulent (NV)
100
90
80
70
60
50

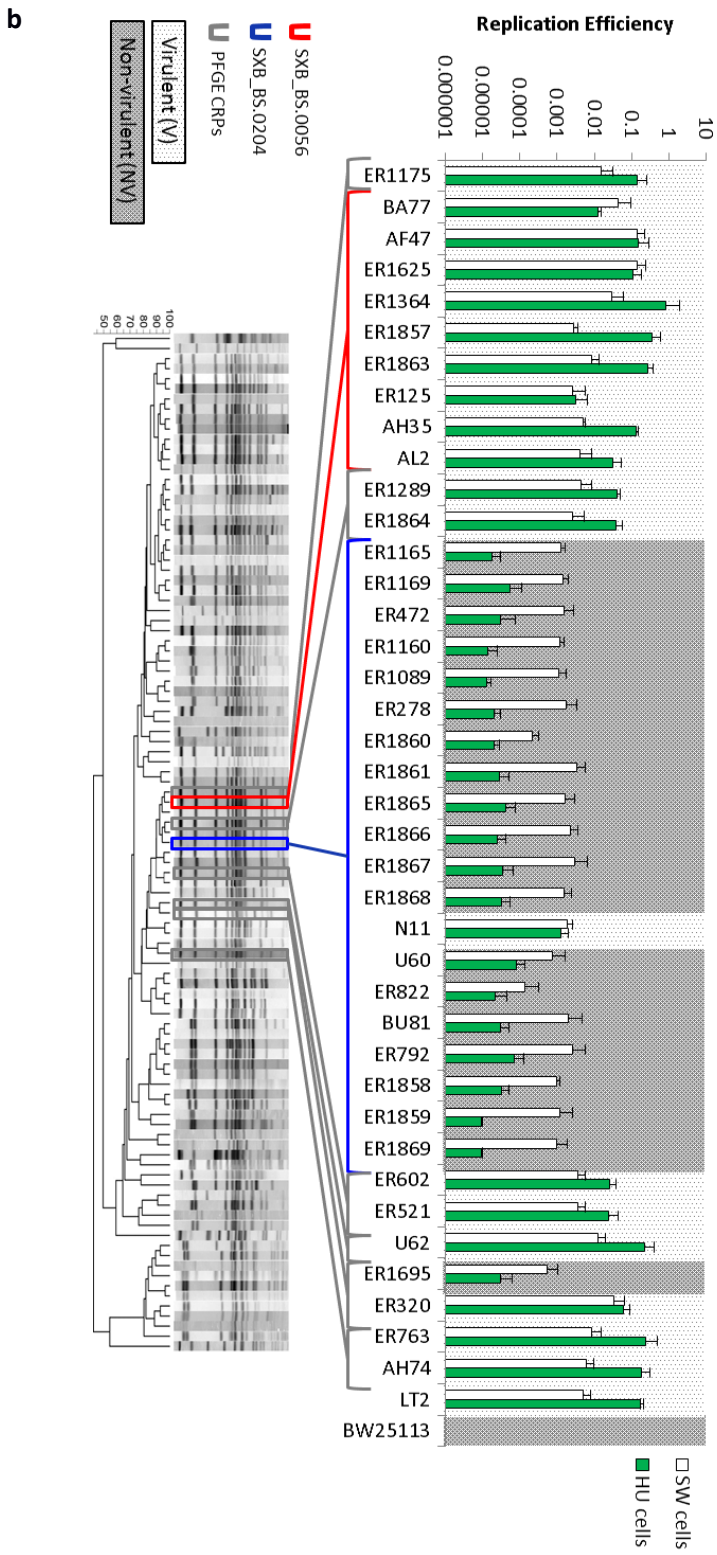


Figure 12: *S.* Derby invasion (a) and replication (b) assays. The invasion efficiency was calculated as mean of CFUs recovered after 1h of infection divided by CFU of the inoculum. The replication efficiency was calculated as mean of CFUs recovered after 1h of infection and 20h of intracellular replication divided by CFU of the inoculum. Error bars correspond to standard deviations of each mean.

3.3. *A*ntibiotic resistance of *S*. Derby isolates

The 39 *S.* Derby isolates selected for invasion and replication assays were screened for resistance to the following antibiotics: Gentamycin (GEN), Ampicillin (AMP), Tetracycline (TET), Kanamycin (KAN), Chloramphenicol (CHL), Spectinomycin (SPC) and Erythromycin (ERY). For each antibiotic, two different concentrations were tested: the epidemiological cutoff value indicated by EUCAST (not available for ERY), and the common working concentrations used for mutants selection, in order to identify selectable markers to use for successive transformation experiments. Gentamycin is included in the screening even to verify absence of resistance to this antibiotic, used in *in vitro* cell lines infection assays.

Results are shown in Table 10. All tested isolates are sensitive to Ampicillin at both tested concentrations, expect for ER763 and AH74, both belonging to SXB_PR.0621 profile (column 1). All isolates showed sensitivity to both tested concentrations of GEN and CHL (columns 2 and 5). SXB_BS.0204, SXB_BS.1274, SXB_PR.0071 and SXB_PR.0787 isolates and ER1864 are sensitive to both TET concentrations and the same isolates, but ER1864, are sensitive to the higher concentration used for SPC. By contrast, SXB_BS.0056 isolates, as well as those with SXB_PR.0119, SXB_PR.0250, SXB_PR.0621, SXB_PR.1018, SXB_PR.1259 profiles and ER1289, are resistant to TET and to the higher SPC concentration. All isolates, except for ER1625, ER822 and SXB_PR.0071 isolates, are resistant only to the lower KAN concentration. No isolates showed sensibility to ERY.

Table 10: Results of antibiotic resistance screening. For each antibiotic the tested concentrations are reported: the epidemiological cutoff value (first column) and the working concentration for mutants selection (second column). Concentrations were expressed in µg/ml. S is reported if the isolate is sensitive to the antibiotic, R if it is resistant.

		Ampicillin (AMP)		Gentamycin (GEN)		Tetracycline (TET)		Kanamycin (KAN)		Chloramphenicol (CHL)		Spectinomycin (SPC)		Erythromycin (ERY)
Strains	**PFGE Genotype**	**8**	**100**	**2**	**15**	**8**	**20**	**2**	**50**	**16**	**20**	**2**	**30**	**20**
BA77	SXB_BS.0056	S	S	S	S	R	R	R	S	S	S	R	R	R
AF47	SXB_BS.0056	S	S	S	S	R	R	R	S	S	S	R	R	R
ER1625	SXB_BS.0056	S	S	S	S	R	R	S	S	S	S	R	R	R
ER1364	SXB_BS.0056	S	S	S	S	R	R	R	S	S	S	R	R	R
ER1857	SXB_BS.0056	S	S	S	S	R	R	R	S	S	S	R	R	R
ER1863	SXB_BS.0056	S	S	S	S	R	R	R	S	S	S	R	R	R
ER125	SXB_BS.0056	S	S	S	S	R	R	R	S	S	S	R	R	R
AH35	SXB_BS.0056	S	S	S	S	R	R	R	S	S	S	R	R	R
AL2	SXB_BS.0056	S	S	S	S	R	R	R	S	S	S	R	R	R
ER1165	SXB_BS.0204	S	S	S	S	S	S	R	S	S	S	R	S	R
ER1169	SXB_BS.0204	S	S	S	S	S	S	R	S	S	S	R	S	R
ER472	SXB_BS.0204	S	S	S	S	S	S	R	S	S	S	R	S	R
ER1160	SXB_BS.0204	S	S	S	S	S	S	R	S	S	S	R	S	R
ER1089	SXB_BS.0204	S	S	S	S	S	S	R	S	S	S	R	S	R
ER278	SXB_BS.0204	S	S	S	S	S	S	R	S	S	S	R	S	R
ER1860	SXB_BS.0204	S	S	S	S	S	S	R	S	S	S	R	S	R
ER1861	SXB_BS.0204	S	S	S	S	S	S	R	S	S	S	R	S	R
ER1865	SXB_BS.0204	S	S	S	S	S	S	R	S	S	S	R	S	R
ER1866	SXB_BS.0204	S	S	S	S	S	S	R	S	S	S	R	S	R
ER1867	SXB_BS.0204	S	S	S	S	S	S	R	S	S	S	R	S	R
ER1868	SXB_BS.0204	S	S	S	S	S	S	R	S	S	S	R	S	R
N11	SXB_BS.0204	S	S	S	S	S	S	R	S	S	S	R	S	R
U60	SXB_BS.0204	S	S	S	S	S	S	R	S	S	S	R	S	R
ER822	SXB_BS.0204	S	S	S	S	S	S	S	S	S	S	R	S	R
ER1858	SXB_BS.0204	S	S	S	S	S	S	R	S	S	S	R	S	R
ER1859	SXB_BS.0204	S	S	S	S	S	S	R	S	S	S	R	S	R
ER1869	SXB_BS.0204	S	S	S	S	S	S	R	S	S	S	R	S	R
BU81	SXB_BS.0204	S	S	S	S	S	S	R	S	S	S	R	S	R
AT39	SXB_PR.0071	S	S	S	S	S	S	S	S	S	S	R	S	R
AD45	SXB_PR.0071	S	S	S	S	S	S	S	S	S	S	R	S	R
ER521	SXB_PR.0119	S	S	S	S	R	R	R	S	S	S	R	R	R
U62	SXB_PR.0250	S	S	S	S	R	R	R	S	S	S	R	R	R
ER763	SXB_PR.0621	R	R	S	S	R	R	R	S	S	S	R	R	R
AH74	SXB_PR.0621	R	R	S	S	R	R	R	S	S	S	R	R	R
ER1695	SXB_PR.0787	S	S	S	S	S	S	R	S	S	S	R	S	R
ER320	SXB_PR.0787	S	S	S	S	S	S	R	S	S	S	R	S	R
ER602	SXB_PR.1018	S	S	S	S	R	R	R	S	S	S	R	R	R
ER1175	SXB_PR.1259	S	S	S	S	R	R	R	S	S	S	R	R	R
ER1289	SXB_PR.1274	S	S	S	S	R	R	R	S	S	S	R	R	R
ER1864	SXB_PR.1274	S	S	S	S	S	S	R	S	S	S	R	R	R

3.4. Virulent and non-virulent isolates form distinct phylogenetic clades

To evaluate the phylogenetic relationship between V and NV isolates, whole genomes of 22 isolates with different virulence phenotype were sequenced on an Illumina MiSeq System and core SNPs were extracted to perform a phylogenetic analysis. Other 23 *S*. Derby isolates with PFGE profiles not closely related to those tested in invasion and replication assays were added to the analysis, in order to determine relationship of the isolates of interest in a broader context of *S*. Derby. A maximum-likelihood tree was constructed based on 7,133 SNPs. AT39, ER386, Z57, AD45 H58 and I18 isolates (SXB_PR.0071 pulsotype), distantly related to all the other isolates (distinguished from the other isolates by a mean of 5960 SNPs), were included in the analysis to identify the root of the tree.

Excluding SXB_PR.0071 isolates, the tree is divided in three clusters (Fig. 13). All SXB_BS.0204 isolates sequenced, including N11 which showed higher invasion and replication efficiencies in human cells compared to the other isolates with the same pulsotype, cluster separately, forming a distinct clade. At least 164 and up to 258 SNPs distinguished NV isolates from those outside this clade. They differed from each other by between 13 and 89 SNPs or between 13 and 54 SNPs, respectively including or excluding N11 from the analysis. N11 differentiates from the other non-virulent isolates for ∼80 SNPs. All V isolates, excluded ER320, were in a single clade and differed from each other for ∼190 SNPs. ER1175 is the virulent isolate belonging to this clade more distant to NV isolates, with an average of 210 SNPs which differentiate it from non-virulent isolates. ER320 is located outside the V cluster, with about 250 SNPs which differentiates it from both the V and NV isolates cluster.

The 22 isolates tested in invasion and replication assays belong to the MLST profile ST40, already described as associated with swine (73).

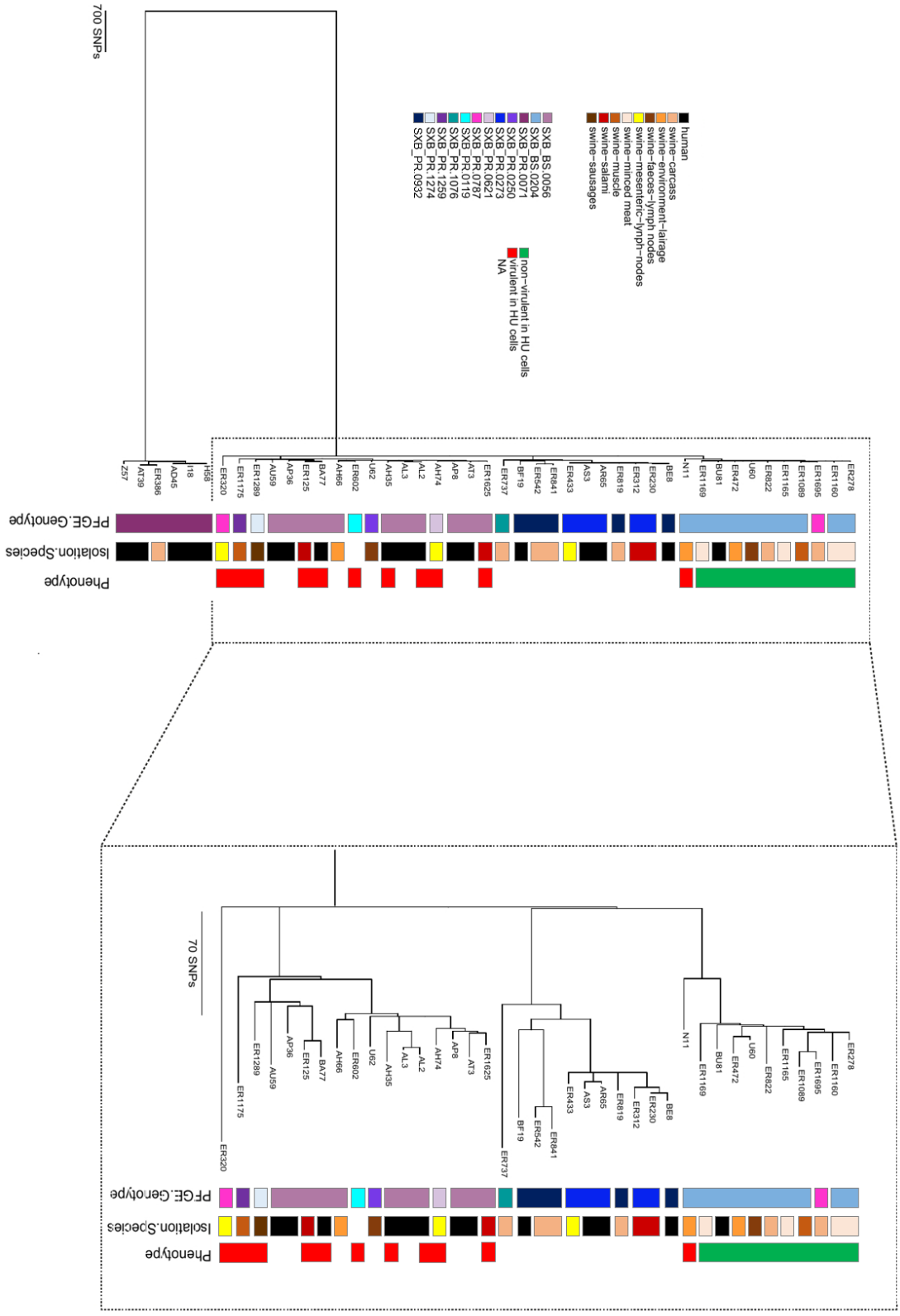


Figure 13: *S*. Derby SNP-based phylogenetic tree. Colors in the first column indicate the PFGE genotype, in the second column the source of isolation and in the last column the virulence phenotype in *in vitro* cell lines infection assays.

3.5. Detection of exclusive genes in virulent isolates and determination of their role in virulence

Host adaptation is associated with a number of genomic signatures, including genomic degradation of generalist genes and gene acquisition by lateral gene transfer (51). An orthology analysis was thus performed to detect genes exclusively present in V isolates, potentially involved in their higher invasion and replication efficiencies in host cells compared to NV isolates. The 22 isolates tested in cell lines infection assays and sequenced were included in the analysis. The pipeline used includes *de novo* assembly of reads for each isolate, protein-coding genes prediction on assembled genomes and identification of classes of orthologous genes. The presence/absence gene matrix obtained was analyzed to assess which genes are unique of V isolates.

The analysis showed that there are no genes present in all V isolates and absent in all NV isolates.

N11 was thus analyzed separately, because of the contrast between its close phylogenetic relationship with NV isolates, and its virulent phenotype in human cells. This isolate was found to carry 30 genes located on an IncFII plasmid (Fig. 14). This plasmid, found only in the N11 isolate, carries genes for conjugal transfer, plasmid replication and stability, DNA modification and regulation and for colicins production and immunity. 22 hypothetical proteins and 3 proteins with unknown function were also annotated on the plasmid sequence. No known virulence genes were found. The analysis was then repeated by eliminating from the group of V isolates both N11 and ER320. ER320 was excluded because, even if it showed a virulence phenotype, the phylogenetic analysis revealed that it is distant from the other V isolates, suggesting different evolutionary paths. The reduced group of V isolates share 69 genes absent in NV isolates. 57 genes are located on the chromosome, composing the *Salmonella* genomic island 1 (SGI-1) and other 6 different regions, called B1-B6. The other 12 genes are located in a region assembled as distinct contig in all V isolates. As BLASTN results showed that the entire region is present in several *Salmonella*

plasmids, it was considered as an hypothetical plasmid region. Genes located in SGI-1, B1-B6 and the hypothetical plasmid region are showed in Fig. 14 and are listed in Table 11. SGI-1 is a genomic island known to be conjugally transferred and integrated into the chromosome in a site-specific manner (120). It carries a complex class 1 integron containing an antibiotic resistance gene cluster identified in several *Salmonella enterica* serovars. In V isolates, the integron carries resistance genes for spectinomycin, sulphonamides and ethidium bromide. Accordingly, V isolates were seen to be resistant to spectinomycin, at both tested concentrations (Table 10). B1, 15,663 bp long, includes phage genes, as well as genes encoding for conjugative transfer proteins. Six hypothetical proteins are also annotated on this region. B2-B5 regions include from 5 to only one gene. They all mapped to prophages inserted in V isolates genomes: B2 belongs to PHAGE_Shigel_SfII_NC_021857, whereas B3, B4 and B5 mapped to PHAGE_Salmon_118970_sal3_NC_031940. The hypothetical plasmid region includes genes for resistance to mercury and tetracycline. Resistance to tetracycline was observed in V isolates in the screening for antibiotic resistance (Table 10).

N11 (N11 WT) was subjected to plasmid curing to lose its exclusive plasmid. Each B1-B6 region and SGI-1 were individually deleted in the virulent isolate ER1175 (ER1175 WT). ER1175 WT was also plasmid cured for the postulated plasmid resulting in loss of hypothetical plasmid region. The obtained mutants are listed in Table 3. Each mutant was tested for virulence in invasion and replication assays in both human and swine cells. Results are expressed as relative invasion and replication, i.e. invasion and replication efficiencies of each mutant normalized to the respective wild type isolate (Fig. 16). Mutants and plasmid-cured strains showed no differences in invasion and replication compared to the respective wild type isolates in both cell lines ($p > 0.05$). All mutants remained fully virulent, therefore the multi-host virulence observed is not caused by genes or plasmids exclusive of V isolates.

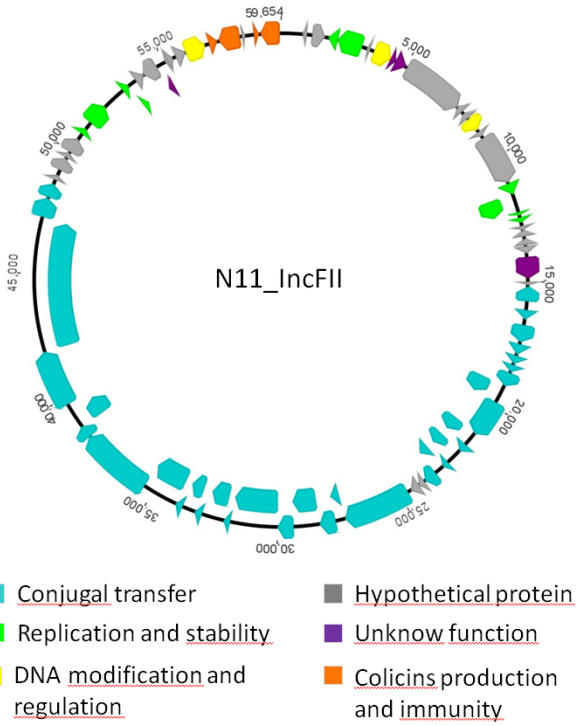


Figure 14: Map of IncFII plasmid found in N11. Genes are differentially colored based on their function.

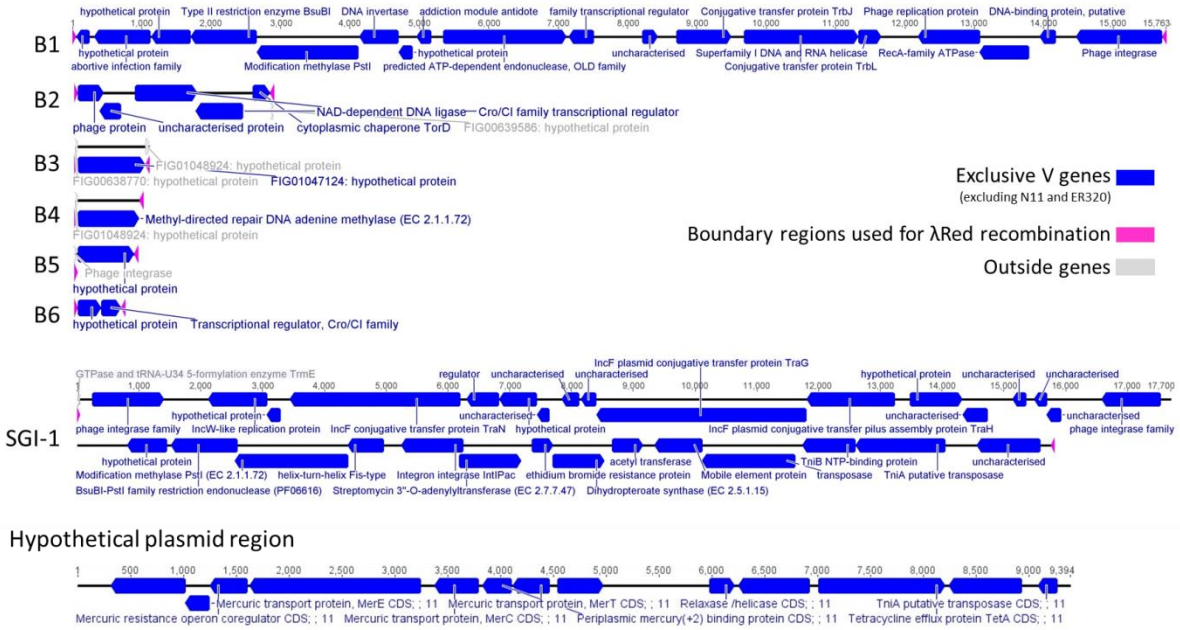


Figure 15: Maps of B1-B6 regions, SGI-1 and the hypothetical plasmid region found in V isolates (excluding N11 and ER320). Exclusive genes are colored in blue. the boundary regions used for λRed recombination in pink and genes outside the identified regions in grey.

Table 11: List of exclusive genes of V isolates located in SGI-1, B1-B6 regions and hypothetical plasmid region. Table reports the length of each region and, for each CDS, its annotation, length and direction.

B1

total length: 15663 bp

CDS Name	lenght	direction
Putative addiction module antidote protein CDS	198	reverse
hypothetical protein CDS	813	reverse
hypothetical protein CDS	567	reverse
Type II restriction enzyme BsuBI (EC 3.1.21.4) CDS	957	reverse
Modification methylase PstI (EC 2.1.1.72) CDS	1,470	reverse
DNA invertase CDS	571	reverse
hypothetical protein CDS	204	reverse
hypothetical protein CDS	204	reverse
predicted ATP-dependent endonuclease, OLD family CDS	1,773	forward
Conjugative transfer protein TraJ CDS	360	reverse
Conjugative transfer protein TraC CDS	210	forward
hypothetical protein CDS	225	forward
Conjugative transfer protein TrbJ CDS	977	forward
Conjugative transfer protein TrbL CDS	1,644	forward
FIG00451338: hypothetical protein CDS	327	forward
Phage replication protein CDS	900	reverse
RecA-family ATPase CDS	726	reverse
DNA-binding protein, putative CDS	231	reverse
Phage integrase CDS	1,251	reverse

B2

total length: 2800 bp

CDS Name	lenght	direction
phage protein CDS	378	forward
uncharacterized protein CDS	309	reverse
NAD-dependent DNA ligase CDS	891	forward
Cro/Cl family transcriptional regulator CDS	687	reverse
cytoplasmic chaperone TorD CDS	249	forward

B3

total length: 975 bp

CDS Name	lenght	direction
hypothetical protein CDS	975	forward

B4

total length: 882 bp

CDS Name	lenght	direction
Methyl-directed repair DNA adenine methylase (EC 2.1.1.72) CDS	882	forward

B5

total length: 813 bp

CDS Name	lenght	direction
hypothetical protein	813	forward

B6

total length: 620 bp

CDS Name	lenght	direction
Putative DNA binding protein or antitoxin igA-2 CDS	333	forward
Conserved hypothetical protein or toxin igB-2	285	forward

SGI-1

total length: 33416 bp

CDS Name	lenght	direction
phage integrase family	1,158	forward
IncW-like replication protein	954	reverse
hypothetical protein	228	reverse
IncF conjugative transfer protein TraN	2,760	reverse
hypothetical protein	534	reverse
hypothetical protein	612	reverse
hypothetical protein	210	reverse
hypothetical protein	291	reverse
hypothetical protein	255	reverse
IncF plasmid conjugative transfer protein TraG	3,405	reverse
IncF plasmid conjugative transfer pilus assembly protein TraH	1,425	reverse
hypothetical protein	858	forward
hypothetical protein	410	reverse
hypothetical protein	219	reverse
hypothetical protein	213	reverse
hypothetical protein	243	reverse
phage integrase family	966	reverse
hypothetical protein	645	reverse
BsuBI-PstI family restriction endonuclease (PF06616)	1,077	reverse
Modification methylase PstI (EC 2.1.1.72)	1,830	reverse
helix-turn-helix Fis-type	588	reverse
Integron integrase IntIPac	1,014	reverse
Streptomycin 3''-O-adenylyltransferase aadA2 (EC 2.7.7.47)	1,008	forward

ethidium bromide resistance protein qacEdelta1	348	forward
Dihydropteroate synthase sul1 (EC 2.5.1.15)	840	forward
acetyl transferase	501	forward
Mobile element protein	783	reverse
transposase	1,515	reverse
TniB NTP-binding protein	861	reverse
TniA putative transposase	1,464	reverse
hypothetical protein	1,035	reverse

Hypothetical plasmid region

total length: 9394bp

CDS Name	**lenght**	**direction**
Tn21 protein of unknown function Urf2 CDS	708	reverse
Mercuric transport protein, MerE CDS	237	reverse
Mercuric resistance operon coregulator CDS	363	reverse
Mercuric ion reductase (EC 1.16.1.1) CDS	1,623	reverse
Mercuric transport protein, MerC CDS	423	reverse
Periplasmic mercury(+2) binding protein CDS	276	reverse
Mercuric transport protein, MerT CDS	351	reverse
Mercuric resistance operon regulatory protein CDS	435	forward
Relaxase /helicase CDS	243	forward
Transcriptional regulator, TetR family CDS	678	reverse
Tetracycline efflux protein TetA CDS	1,200	forward
Permease of the drug/metabolite transporter (DMT) superfamily CDS	699	reverse
TniA putative transposase CDS	192	reverse

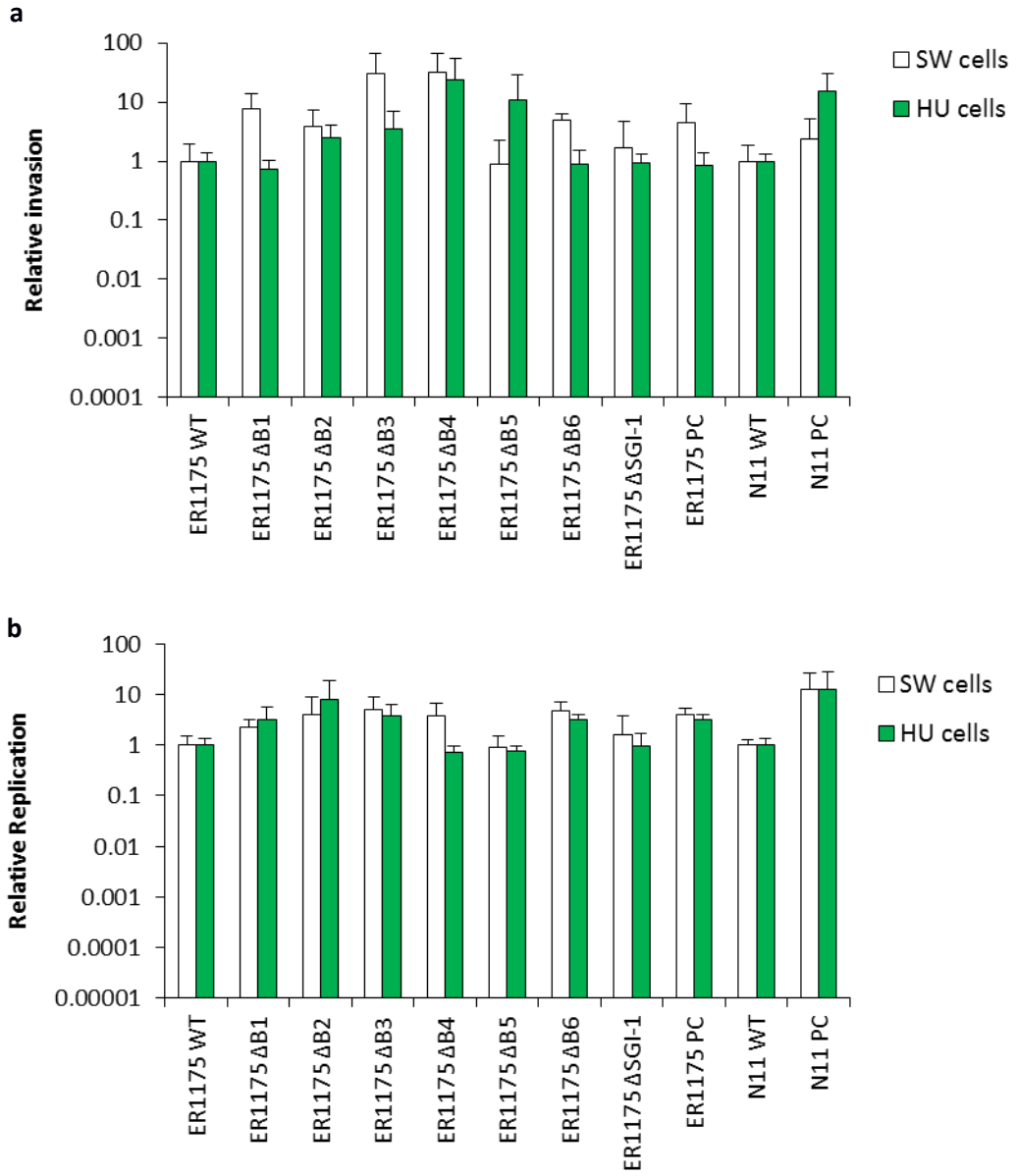


Figure 16: invasion (a) and replication (b) assays on ER1175 WT, N11 WT and their mutants deleted for their unique genes and plasmids. Results are expressed as relative invasion and replication, i.e. the invasion and replication efficiencies of each strain normalized on that of respective WT. Error bars corresponds to standard deviations of each mean.

3.6. Selection of nsSNPs discriminating non-virulent isolates from virulent isolates: one SNP in *hilD* is responsible for differences in virulence phenotype in human cells

The ability of *Salmonella* to colonize different hosts may depend not only on the presence of specific genes, but also on the allelic variation within genes (55, 56). To determine if specific allelic variants are responsible for the observed differences in virulence between V and NV isolates, SNPs that discriminate the two groups were selected. A total of 30 SNPs were detected (Table 12): 2 SNPs are located in intergenic regions, 10 are synonymous SNPs, and 18 are non-synonymous SNPs. Among non-synonymous SNPs, 6 missense conservative mutations, 2 non-sense mutations, and 10 missense non-conservative mutations were found. The 2 non-sense mutations are located in the *ydiV* and *yhaK* genes. *ydiV* encodes for a repressor of flagellar class II operons, decreasing motility in nutrient-poor medium (121). It is also known to be required for resistance to host phagocyte oxidase, for suppression of flagellin-mediated killing of macrophages, and for promoting resistance to hydrogen peroxide (122). *yhaK* encodes for a protein of unknown function, found in low abundance in the cytosol and shown to be a good marker for monitoring oxidative stresses in *E. coli* (123). Both non-sense mutations are located in the middle of the genes (*ydiV* = c.367C>T p.Gln123*; *yhaK* = c. 354G>A p.Trp118*), resulting in a truncated product and likely in significant functional impacts.

One missense non-conservative mutation is located into the *hilD* gene (*hilD* = c. 872G>A p.Cys118Tyr). *hilD* encodes for the major activator of SPI-1, known to be essential for invasion of human epithelial cells. SPI-1 encodes for the type three secretion system (T3SS), a ‘‘molecular syringe’’ that injects effector proteins into the host cell cytosol, promoting actin cytoskeletal rearrangements and Salmonella engulfment (29, 124). HilD is known to regulate also other

virulence genes and pathogenicity islands (103). It can induce the expression of SPI-2 genes (125), needed mainly for *Salmonella* survival and replication within macrophages and consequent establishment of systemic disease, of SPI-4 genes (126), involved in the adhesion of *Salmonella* to polarized epithelial cells, and of flagellar genes (127), known to be responsible for motility but even for a T3SS-1 independent mechanism of invasion of Peyer's patches (36). V and NV isolates carry two different *hilD* allelic variants: the V allele (*hilD_v*) has a codon TGT coding for cysteine whereas the NV allele (*hilD_nv*) has a codon TAT coding for tyrosine at position 291. This mutation is immediately downstream of one of the amino acids predicted to make base-specific contact with DNA (128). Furthermore, the fold stability change (ΔΔG) of HilD encoded by *hilD_nv* resulted in a negative value (-3.09), indicating a strong destabilizing effect on the protein fold caused by the missense mutation.

The two non-sense mutations and the missense non-conservative mutation found in *hilD* were thus investigated for their role in virulence.

Concerning the non-sense mutations found in the *ydiV* and *yhaK* genes in NV isolates, two mutants of ER1175 were generated: the V allelic variants of *ydiV* and *yhaK* where replaced by the NV allelic variants, to obtain the strains ER1175::*ydiV_nv* and ER1175::*yhaK_nv*, respectively. Mutants were tested for virulence by invasion and replication assays in INT-407 and IPEC-J2 cell lines. We found no differences in invasion and replication efficiency compared to ER1175 WT, both in human and swine cells (Fig. 17, a and b respectively, $p>0.05$). Therefore, differences in virulence between V and NV isolates observed in invasion and replication assays are not caused by these two SNPs.

Table 12: SNPs which discriminate NV from V isolates. The table reports, for each genic SNP, the gene, the product name, the position inside the gene, the codons and the corresponding amino acids present in V and in NV isolates. For intergenic SNPs, only the substitution is showed.

					V isolates		NV isolates	
	Type of SNP	**Gene**	**Product name**	**SNP position in CDS**	**codon**	**amino acid**	**codon**	**amino acid**
1	Genic	*panE*	2-dehydropantoate reductase	592/912	CGC	Arg	TGC	Cys
2	Genic	*helD*	Helicase IV	1726/2055	GCG	Ala	GAG	Glu
3	Genic	*astC*	Succinylornithine transaminase	1125/1227	TTT	Phe	TTG	Leu
4	Genic	*ydiV*	Anti-FlhC(2)FlhD(4) factor	367/714	CAG	Gln	TAG	STOP
5	Genic		RpoE-regulated lipoprotein	116/576	ACA	Thr	ATA	Ile
6	Genic	*gabD*	Succinate-semialdehyde dehydrogenase [NADP(+)]	932/1449	CAG	Gln	CTG	Leu
7	Genic	*hilD*	Transcriptional regulator	872/930	TGT	Cys	TAT	Tyr
8	Genic	*lgt*	Prolipoprotein diacylglyceryl transferase	829/876	ATG	Met	GTG	Val
9	Genic	*yhaK*	Pirin-like protein	354/702	TGG	Trp	TGA	STOP
10	Genic	*hemN*	Oxygen-independent coproporphyrinogen-III oxidase	1172/1374	CAG	Gln	CTG	Leu
11	Genic	*coaA*	Pantothenate kinase	529/927	GAA	Glu	AAA	Lys
12	Genic	*zraR*	Transcriptional regulatory protein	550/1326	GCC	Ala	TCC	Ser
13	Genic		Hypothetical protein	167/438	CCG	Pro	CTG	Leu
14	Genic		Lipoprotein	671/720	CCG	Pro	CTG	Leu
15	Genic	*dppA*	Dipeptide-binding protein	263/1608	GGT	Gly	GTT	Val
16	Genic	*dlgD*	2,3-diketo-L-gulonate reductase	329/999	GCG	Ala	GTG	Val
17	Genic	*ybaN*	Inner membrane protein	209/378	CCG	Pro	CTG	Leu
18	Genic	*fhuB*	Ferrichrome transport protein	1843/2058	GTT	Val	ATT	Ile
19	Genic	*chiA*	Chitinase A precursor	687/1764	GAC	Asp	GAT	Asp
20	Genic		ParB-like nuclease domain protein	192/618	GTT	Val	GTA	Val
21	Genic	*cheA*	Chemotaxin protein	159/2016	GGC	Gly	GGT	Gly
22	Genic	*zwf*	Glucose-6-phosphate 1-dehydrogenase	513/1476	CAG	Gln	CAA	Gln
23	Genic	*rcsC*	Sensor kinase protein	649/2763	CAC	His	CAT	His
24	Genic	*pnbA*	Para-nitrobenzyl esterase	306/1509	CCG	Pro	CCA	Pro
25	Genic		Hypothetical protein	1137/1344	ATC	Ile	ATT	Ile
26	Genic	*rpsU*	30S ribosomal protein S21	90/216	GCG	Ala	GCA	Ala
27	Genic	*nfi*	Endonuclease V	312/672	GGT	Gly	GGC	Gly

28	Genic	Phosphoglycerate mutase	377/648	AGT	Ser	AGC	Ser
29	Intergenic			C		T	
30	Intergenic			G		T	

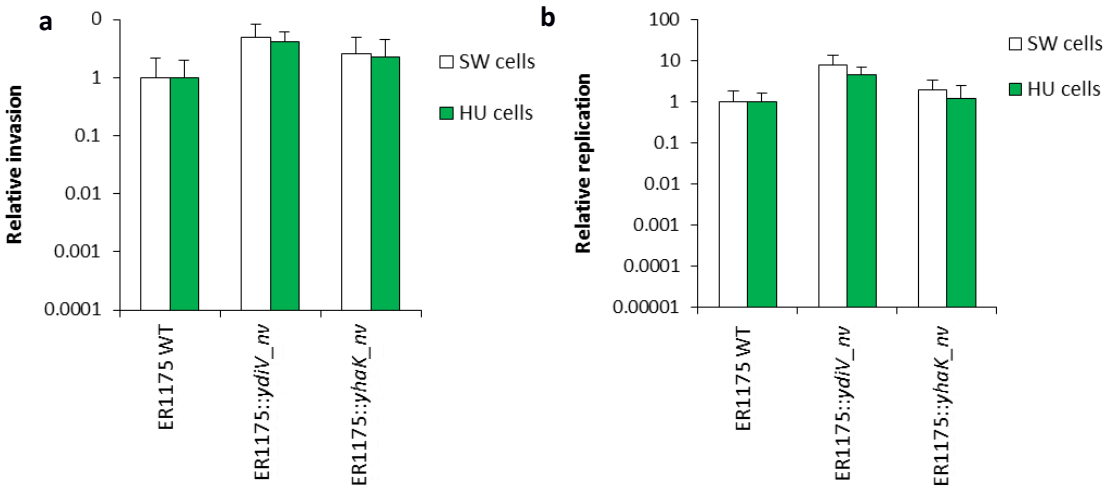


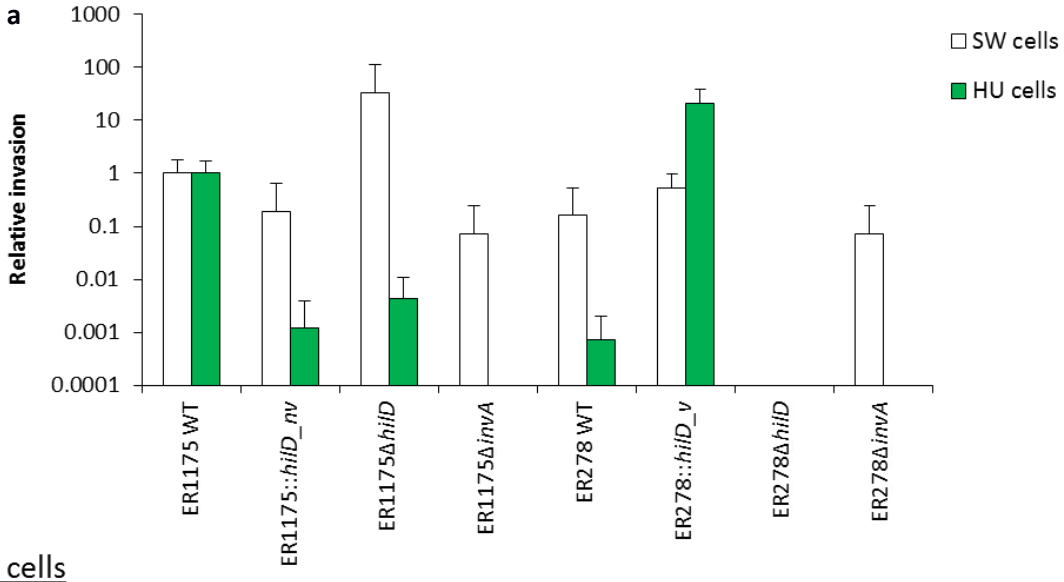
Figure 17: invasion (a) and replication (b) assays on ER1175 WT, ER1175::*ydiV_nv* and ER1175::*yhaK_nv*. Results are expressed as relative invasion and replication, i.e. the invasion and replication efficiencies of each strain normalized on that of respective WT. Error bars corresponds to standard deviations of each mean.

To assess the role of *hilD* allelic variants in the virulence phenotypes observed, *hilD_v* in ER1175 WT was replaced by recombination with the NV allele to obtain the mutant ER1175::*hilD_nv*. Results on human cells (Fig 18, a and b, green bars) shows that ER1175::*hilD_nv* has the same low invasion and replication efficiencies in human cells of ER278 (ER278 WT), a NV isolate (p>0.05 for both invasion and replication). Accordingly, *hilD_nv* in ER278 WT was replaced with the V allele to generate the mutant ER278::*hilD_v*, which shows even higher invasion and replication efficiencies compared to ER1175 WT (p=0.009 relative to both invasion and replication assays). The two *hilD* allelic variants are thus responsible for different invasion and replication efficiencies of V and NV isolates in human cells. We thus deleted the *hilD* gene in ER1175 WT and ER278 WT (ER1175Δ*hilD* and ER278Δ*hilD*) to assess if the *hilD_nv* allele confers the same virulence phenotype of deletion of the entire gene. We found that ER1175Δ*hilD* invaded and replicated in human cells at the same level of ER278 WT (p>0.05), whereas neither invasion nor replication were detected for ER278Δ*hilD*. These results suggest that the *hilD_nv* allele produces a non-functional protein.

HilD, together with HilC and RtsA, constitutes a feed-forward loop that controls expression of the SPI-1 regulator HilA. Null mutations in *hilD* decrease expression of *hilA* ~10-fold under SPI-1 inducing conditions, whereas *hilC* or *rtsA* mutations decrease expression of *hilA* approximately two-fold (129-131). *hilD* is considered the major regulator of *hilA*, and therefore of SPI-1, which encodes for T3SS-1. To evaluate if loss of a functional T3SS-1 confers the same loss of invasion and replication efficiencies in human cells caused by the presence of *hilD_nv* or *hilD* deletion, we deleted the *invA* gene, which encodes for a main structural component of the T3SS-1, in both strains (ER1175Δ*invA* and ER278Δ*invA*). ER1175Δ*invA* invaded and replicated in human cells at the same low level of ER278 WT (p>0.05), whereas for ER278Δ*invA*, like ER278Δ*hilD*, neither invasion nor replication were detected. The phenotypes observed in human cells indicate that carrying *hilD_nv* elicits the same (for ER1175) or a similar (for ER278) virulence phenotype of loss of a

functional T3SS-1, suggesting that the *hilD_nv* product is not able to activate the expression of *hilA* and, in turn, of all SPI-1 genes.

A different scenario was observed in swine cells (Fig 18, a and b, white bars). ER1175::*hilD_nv* and ER1175Δ*hilD* showed no differences in invasion and replication efficiencies compared to ER1175 WT ($p>0.05$), whereas ER1175Δ*invA* infected at a significantly lower extent ($p<0.001$), comparable to that of ER278 WT ($p>0.05$). ER278Δ*hilD* and ER278Δ*invA* invaded and replicated in swine cells at the same level of ER278 WT ($p>0.05$), significantly lower than that of ER278::*hilD_v*, which is comparable to ER1175 WT. *hilD* appeared to be dispensable for infection of swine cells in ER1175, as only the deletion of *invA*, but not the deletion of *hilD* or the presence of *hilD_nv*, caused a decrease of invasion and replication efficiencies. By contrast, deletion of *invA* in ER278 did not cause significant changes in the virulence phenotype compared to deletion of *hilD* or the presence of *hilD_nv*. However, the introduction of *hilD_v* provoked a significant increase of invasion and replication efficiencies. Collectively, these results indicate that HilD and SPI-1 may have different roles in infection of human and swine cells.

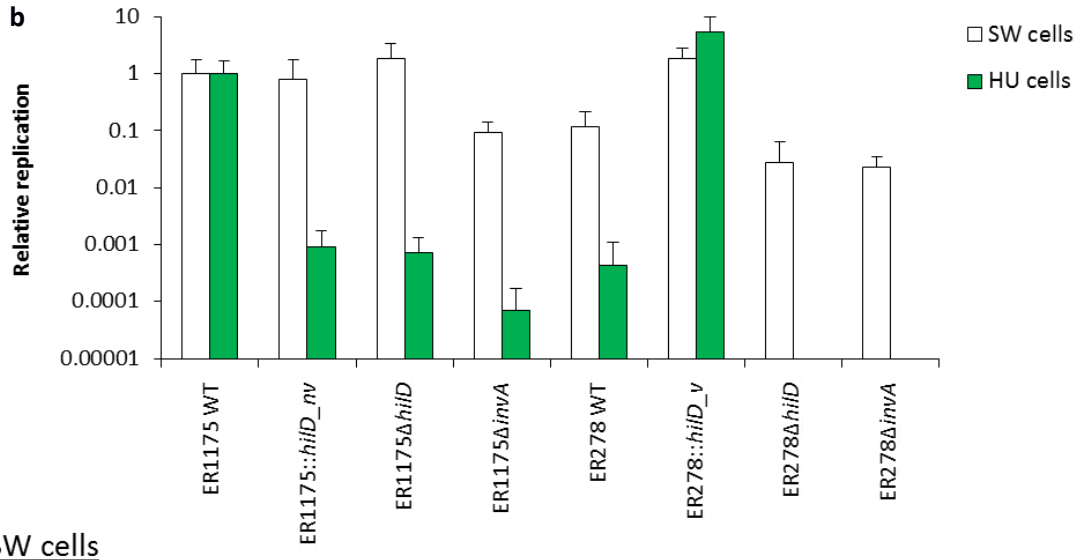


SW cells

vs ER1175 WT	-	< 0.05	n.s.	< 0.001	< 0.01	n.s.	< 0.001	< 0.001
vs ER278 WT	< 0.01	n.s.	< 0.05	n.s.	-	< 0.01	n.s.	n.s.

HU cells

vs ER1175 WT	-	< 0.001	< 0.001	< 0.001	< 0.001	< 0.01	< 0.001	< 0.001
vs ER278 WT	< 0.001	n.s.	n.s.	n.s.	-	< 0.001	< 0.001	< 0.001



SW cells

vs ER1175 WT	-	n.s.	n.s.	< 0.001	< 0.01	n.s.	< 0.001	< 0.001
vs ER278 WT	< 0.01	n.s.	< 0.05	n.s.	-	< 0.01	n.s.	n.s.

HU cells

vs ER1175 WT	-	< 0.001	< 0.001	< 0.001	< 0.001	< 0.01	< 0.001	< 0.001
vs ER278 WT	< 0.001	n.s.	n.s.	n.s.	-	< 0.001	< 0.001	< 0.001

Figure 18: invasion (a) and replication (b) assays on ER1175 WT, ER278 WT and their mutants generated to understand the role of the observed SNP in *hilD*. Results are expressed as relative invasion and replication, i.e. the invasion and replication efficiencies of each strain normalized on that of ER1175 WT. Error bars corresponds to standard deviations of each mean. Tables report *p*-values representing the results of Monte Carlo EM.

3.7. Whole transcriptome analysis to assess differences in genes expression caused by *hilD* allelic variants

The virulence phenotypes of *hilD* and *invA* mutants observed in human cells suggest that *hilD_nv* encodes for a non-functional protein that does not activate the expression of SPI-1 genes. To test this hypothesis, the gene expression profiles of ER1175 WT, ER278 WT, ER1175::*hilD_nv*, ER278::*hilD_v*, ER1175Δ*hilD*, ER278Δ*hilD* strains were compared. Total RNA was extracted from each strain cultured to early stationary phase, because Colgan *et al* (103) showed that, in this particular growth phase, expression of the SPI-1 genes was reduced by an average of 40-fold in a *S*. Typhimurium mutant deleted for *hilD*. RNA samples were used to construct cDNA libraries for RNA-seq. The obtained reads were mapped against ER1175 draft genome, deleted for genes absent in ER278 to avoid false positive, i.e. genes resulting significantly up-regulated in ER1175 strains compared to ER278 strains because ER278 does not carry those genes. For each sample, the number of reads per gene was normalized by size factor. Results of differential expression analysis between strains were summarized in six pair-wise comparisons, represented in MA plots (Fig. 19), i.e. scatter plots illustrating for each gene the log2 fold change on the y-axis and the average of the normalized reads counts on the x-axis. The strain pairs analyzed were indicated above the respective plot. Genes considered significantly up- or down-regulated (with a *p*-value below the threshold, 0.1) in the first strain compared to the second one are represented by red dots.

We found that 51 genes were significantly upregulated in ER1175 WT (carrying *hilD_v*) compared to both ER1175::*hilD_nv* and ER1175Δ*hilD* (Fig. 19, a and b). In ER278::*hilD_v*, 142 and 169 genes were up-regulated, and 149 and 169 genes were down-regulated compared to ER278 WT and ER278Δ*hilD*, respectively (Fig. 19, d and e). No genes were differentially expressed in ER1175::*hilD_nv* compared to ER1175Δ*hilD* as well as in ER278 WT compared to ER278Δ*hilD*,

except for *hilD* itself because it is absent in deleted strains (Fig. 19, c and f). Both ER1175 and ER278 strains carrying the *hilD*_v allele (ER1175 WT and ER278::*hilD_v*) shared 45 up-regulated genes (Fig. 20 and Table 13). This group includes:

- 32 of the 50 SPI-1 encoded genes,
- *rtsA* and *rtsB*, encoded in the same operon outside SPI-1 and involved in SPI-1 and flagellar operons regulation, respectively (132),
- 4 genes (*siiA*, *siiB*, *siiC, siiD*) encoded in one SPI-4 operon: as already mentioned, SPI-4 is known to be regulated by *hilD* (126),
- 2 genes, *sopB* (*sigD*) *and pipC* (*SigE*) encoded in an operon located in SPI-5: they are already known to be regulated by *invF* and *sicA*, in turn activated by hilA (133),
- 3 genes of unknown function: for two of them the locus tag from *S*. Typhimurium LT2 is indicated, the third one is not annotated on LT2 genome.

Overall, these results show that the *hilD_v* allele encodes for a functional transcriptional regulator, able to activate the expression of SPI-1, SPI-4, and SPI-5 genes. Three other genes never associated with *hilD* also appeared to be regulated by this activator. The results confirms that the point mutation in *hilD* found in NV isolates causes a total loss of function of the encoded regulator, as no differences in gene expression were found between both ER1175 and ER278 carrying *hilD_nv* (ER1175::*hilD_nv* and ER278 WT) and the same strains deleted for *hilD*. Interestingly, the replacement of the *hilD_nv* allele with *hilD_v* in ER278 caused changes in expression of about 250 genes other than the only 45 genes in ER1175 WT. This result suggests that HilD could have a role in regulation of other pathways in NV isolates, likely due to NV-isolates or even ER278 specific genetic background.

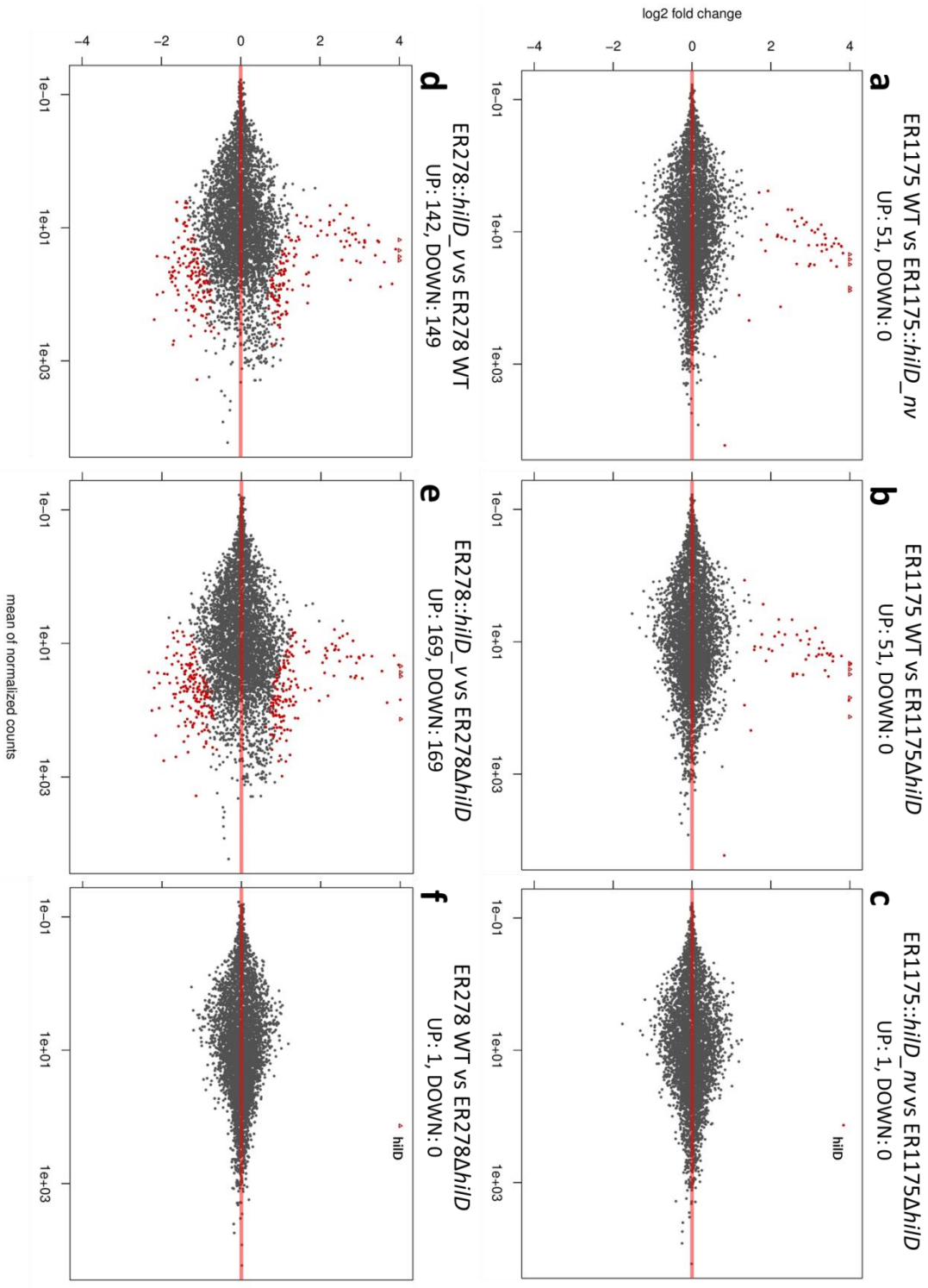


Figure **19**: Whole transcriptome analysis MA-plots. The log2 fold change for a particular comparison is plotted on the y-axis and the average of the counts normalized by size factor is shown on the x-axis. Each gene is represented with a dot. Genes with an adjusted *p*-value below the threshold (here 0.1, the default) are shown in red.

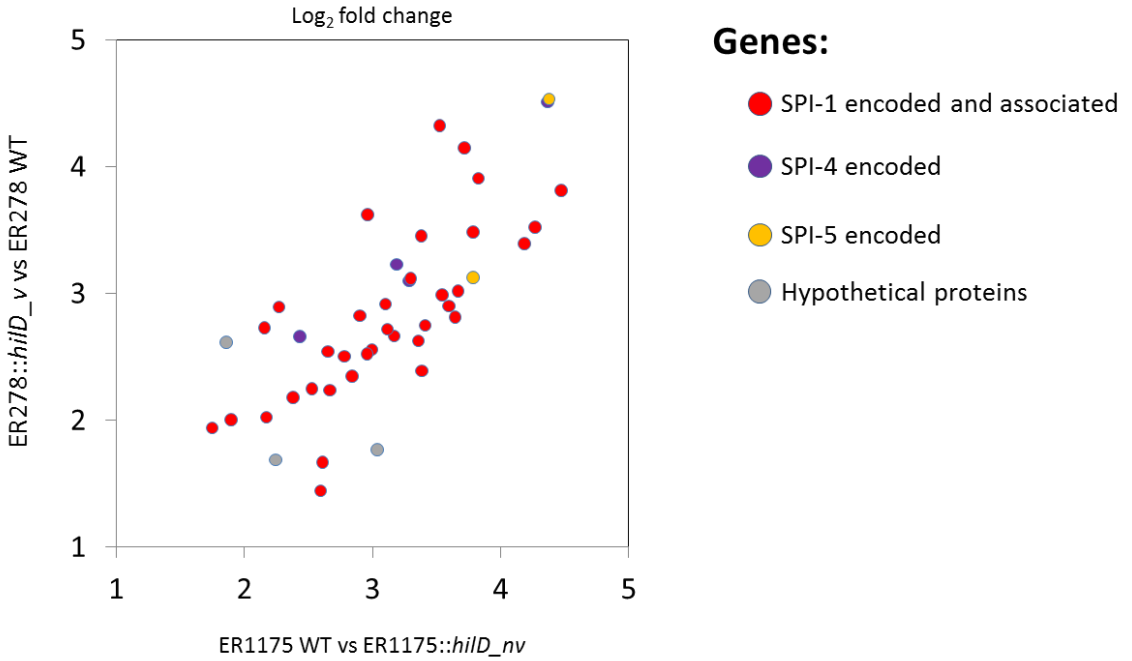


Figure 20: log2 fold change values of common significantly up-regulated genes in ER278::hilD_v vs ER278 WT (y-axis) and ER1175 WT vs ER1175::hilD_nv (x-axis). Each gene is represented with a dot that is colored based on annotation**.**

Table 13: list of genes significantly up-regulated in ER278::hilD_v vs ER278 WT and ER1175 WT vs ER1175::hilD_nv. For each gene the name, the log2 fold change value refer to ER1175 WT and ER278::hilD_nv and SPI reference are reported.

	log2 fold change		
gene	**ER1175 WT vs ER1175::*hilD_nv***	**ER278::*hilD_v* vs ER278WT**	
avrA	1.749984	1.937797	SPI-1 encoded
sprB	3.715781	4.147595	SPI-1 encoded
hilC	3.524141	4.323323	SPI-1 encoded
orgC	2.269207	2.891061	SPI-1 encoded
orgB	2.378921	2.180607	SPI-1 encoded
orgA	2.156421	2.728669	SPI-1 encoded
prgK	2.960942	3.620831	SPI-1 encoded
prgJ	3.100966	2.916724	SPI-1 encoded
prgI	2.997441	2.556516	SPI-1 encoded
prgH	2.653712	2.541493	SPI-1 encoded
hilD	2.243106	1.685257	SPI-1 encoded
hilA	1.858441	2.612191	SPI-1 encoded
iagB	1.895134	2.003297	SPI-1 encoded
sptP	3.411457	2.747942	SPI-1 encoded

sicP	3.646775	2.812114	SPI-1 encoded
STM2880	2.608941	1.667095	SPI-1 encoded
sipA	3.668217	3.016625	SPI-1 encoded
sipD	3.543851	2.985868	SPI-1 encoded
sipC	4.267367	3.522038	SPI-1 encoded
sipB	4.471572	3.81223	SPI-1 encoded
sicA	3.17081	2.664587	SPI-1 encoded
spaP	2.668878	2.234605	SPI-1 encoded
spaO	2.839263	2.34848	SPI-1 encoded
invJ	3.357004	2.626704	SPI-1 encoded
invI	2.901141	2.825272	SPI-1 encoded
invC	3.597131	2.898727	SPI-1 encoded
invB	3.386612	2.388342	SPI-1 encoded
invA	4.185283	3.390911	SPI-1 encoded
invE	3.38096	3.454023	SPI-1 encoded
invG	3.785062	3.481512	SPI-1 encoded
invF	3.826461	3.907764	SPI-1 encoded
invH	2.782307	2.502646	SPI-1 encoded
sopE2	3.188682	3.226866	SPI-1 associated
sopE	3.285716	3.099274	SPI-1 associated
rtsB	2.431495	2.658793	SPI-1 associated
rtsA	4.366995	4.507976	SPI-1 associated
siiD	2.594721	1.444409	SPI-4 encoded
siiC	3.037648	1.768436	SPI-4 encoded
siiB	3.116263	2.715569	SPI-4 encoded
siiA	2.526558	2.248084	SPI-4 encoded
sopB (sigD)	4.37851	4.533076	SPI-5 encoded
pipC (sigE)	3.786647	3.12446	SPI-5 encoded
hypothetical protein CDS	2.171057	2.022193	
hypothetical protein CDS (STM1328)	3.297549	3.118252	
hypothetical protein CDS (STM1239)	2.957699	2.52022	

3.8. Loss of SPI-1 genes expression in non-virulent isolates is verified during human cell infection

Differential expression analysis on RNA extracted during early stationary phase showed that carrying *hilD_nv*, as well as the *hilD* deletion, causes loss of expression of several genes, including many SPI-1 genes. To evaluate if this phenomenon occurs also during infection of human cells, bacterial RNA of ER1175, ER1175::*hilD_nv* and ER1175Δ*hilD* was extracted after 0, 30 and 60 minutes of infection (T0, T30, T60) and qPCRs were performed on *hilD*, other 6 SPI-1 genes and *pagN*. The SPI-1 genes analyzed are 4 transcriptional activators (*hilC*, *rtsA*, *hilA,* and *invF*) and 2 representative genes of the SPI-1 type III secretion system and effector proteins injected into host cells (*invA* and *sipB*), respectively. *pagN* supports adhesion and invasion of epithelial cells, and is known to be regulated by the PhoPQ system, but not by HilD (134, 135). Results for each gene in each condition were expressed as fold change values relative to ER1175 WT at T0, used as control condition (Table 14).

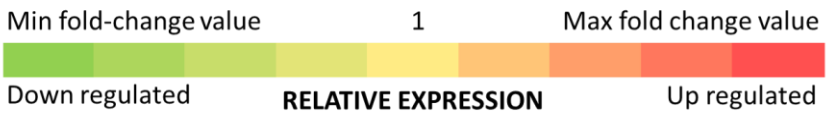
In ER1175 WT, we observed an increased expression of all genes analyzed over time, ranging from 4.67 to 69.52-fold change at T30, and from 4.21 to 72.71-fold change at T60, compared to ER1175 WT – T0. At T0, in both ER1175::*hilD_nv* and ER1175Δ*hilD* the expression of SPI-1 genes was significantly downregulated compared to ER1175 WT, and no increase in genes expression was detected at T30 and T60. Differences in SPI-1 genes expression were therefore verified not only in *in vitro* growth condition, but even during infection of human cells.

The dynamics of *pagN* expression were different from those of SPI-1 genes. *pagN* was upregulated in ER1175::*hilD_nv* and ER1175Δ*hilD* at T30 and T60 compared to the respective T0 and ER1175 WT at T0, but, at T60, its expression was significantly lower in the mutants compared to the wild type strain (p>0.05). Therefore, ER1175::*hilD_nv* and ER1175Δ*hilD* upregulate *pagN* expression

over time, but to a lower extent than ER1175 WT. Intriguingly, the three strains analyzed showed no differences in *pagN* expression when RNA was extracted during early stationary phase for RNA sequencing. These results may suggest that HilD could be indirectly involved in *pagN* regulation, depending on the environmental conditions.

Table 14: Quantitative RT-PCR of ER1175 WT and its mutants during infection of human cells. Values listed are relative fold-change values as compared to ER1175 WT–T0 (control condition). Bold values are significantly different from ER1175 WT-T0.

	ER1175 WT			ER1175::*hilD_nv*			ER1175Δ*hilD*		
	T0	T30	T60	T0	T30	T60	T0	T30	T60
hilD	1	**10.74**	**14.72**	0.67	1.37	**2.24**			
hilA	1	**5.57**	**5.47**	**0.30**	0.84	1.21	0.97	0.92	1.20
hilC	1	**4.67**	**6.09**	**0.13**	**0.03**	**0.08**	**0.26**	**0.05**	**0.05**
rtsA	1	**34.02**	**36.21**	**0.42**	**0.25**	0.57	0.69	0.23	**0.27**
sipB	1	**69.52**	**72.71**	**0.06**	**0.06**	**0.14**	**0.02**	**0.04**	**0.11**
invA	1	**6.10**	4.21	**0.04**	**0.06**	**0.09**	**0.10**	**0.04**	**0.04**
invF	1	**3.89**	**4.50**	**0.09**	**0.09**	**0.10**	**0.16**	**0.14**	**0.06**
pagN	1	**34.60**	**52.39**	**1.65**	**25.88**	**19.28**	2.82	**23.84**	**13.62**



3.9. Assessing differences in *in vivo* host-bacterial interaction caused by *hilD* allelic variants

Experiments with *S*. Typhimurium have shown that SPI-1 plays a key role in eliciting intestinal inflammation in the streptomycin-treated mouse model of colitis (136, 137). A mutant deleted for *invG*, encoding for a structural component of T3SS-1, as well as mutants deleted individually for *sipA*, *sopE* and *sopE2*, encoding for T3SS-1 translocated effectors, cause only mild gut inflammation.

To assess if the downregulation of SPI-1 genes observed in *S*. Derby strains carrying *hilD_nv* causes altered pathogenicity *in vivo*, streptomycin-pretreated mice were infected with either ER1175 WT, ER1175::*hilD_nv*, ER1175Δ*hilD*, or ER278 WT. Three days post infection, mice were humanely sacrificed to analyze the bacterial loads in caecum, spleen and liver, and for the analysis of expression of host pro-inflammatory genes. We found that bacterial loads in the cecum (Fig. 21a) were not significantly different among mice infected with either tested strains ($p>0.05$), except for ER1175Δ*hilD*, which was recovered in higher amount than ER1175::*hilD_nv* ($p=0.008$). In the spleen (Fig. 21b), we observed significantly higher load of ER1175Δ*hilD* than ER1175 WT ($p<0.05$), whereas ER1175 WT, ER278 WT and ER1175::*hilD_nv* colonized the spleen to similar levels ($p>0.05$). A comparable scenario was observed in liver (Fig. 21c): even though it did not reach statistical significance, ER1175Δ*hilD* load was higher than that of ER1175 WT, and no differences were detected among bacterial load of ER1175 WT, ER278 WT and ER1175::*hilD_nv* ($p>0.05$). To evaluate the degree of gut inflammation, total cecal RNA from infected and uninfected mice was extracted, reversely transcribed, and SYBR Green based real-time PCRs were performed on *Cxcl1* and *Nos2* genes. *Cxcl1* encodes for a chemoattractant of neutrophils, cells that migrate into the intestinal lumen upon *Salmonella* host cells invasion and induction of proinflammatory cytokine production.

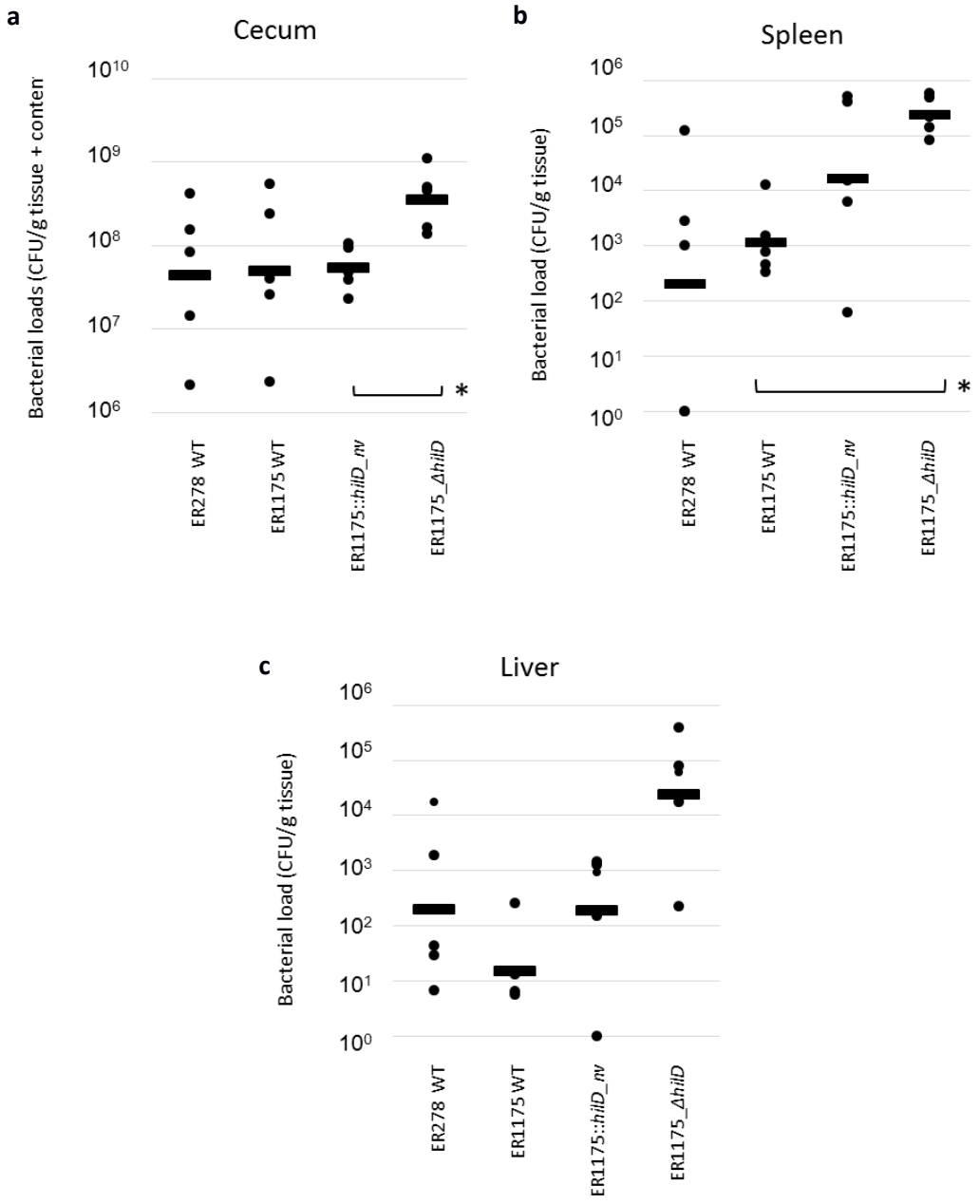


Figure 21: Bacterial loads of each tested strain in cecum, spleen and liver. Each dot represents a mouse. The bars indicate the geometric mean of the bacterial loads. * symbol indicates when *p*-values is < 0.05.

Nos2 encodes for a nitric oxide synthase induced during gut inflammation: the consequent production of nitric oxide (NO) represents an important microbicidal mechanism exploited by the innate immune system. Expression of these genes has been already used to monitor gut inflammation (38).

In our study, we found no significant differences among strains in the expression of both *Cxcl1* and *Nos2* genes (Table 15). These results suggest that *S*. Derby isolates may not able to cause gut inflammation in mice.

Table 15: Quantitative RT-PCR on ER1175 WT, ER1175::*hilD_nv*, ER1175Δ*hilD* and ER278 WT infected and uninfected mice. For each gene fold-change over uninfected control mice, standard errors (in brackets) and *p*-values are reported.

	Nos2		***Cxcl1***	
	Fold change	*p*-value	Fold change	*p*-value
ER278 WT	0.938 (0.199 - 3.482)	0.93	0.471(0.155 - 1.193)	0,177
ER1175 WT	1.486 (0.300 - 5.037)	0.497	0.581(0.142 - 3.585)	0,483
ER1775::*hilD*_nv	2.822 (0.916 - 9.897)	0.128	0.799 (0.340 - 2.107)	0,625
ER1175Δ*hilD*	1.797 (0.609 - 6.053)	0.352	0.644 (0.263 - 1.312)	0,306
Uninfected mice	1 (0.201 – 4.975)	0.952	1 (0,508 - 1,978)	0,96

4. Discussion

The last summary on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks produced by EFSA reported that *Salmonella* is the second cause of food-borne disease in Europe. It is therefore critically important to improve the ability of the surveillance systems to assess the risk of *Salmonella* transmission along the food-chain. The probability of cross-species transmission is directly correlated to host-tropism. *Salmonella* serotypes can vary widely in their ability to cause disease in different host species, ranging from specialists, largely restricted to a single host, to generalists, which exhibit a wide host range. There is also evidence of different host tropism among isolates with different sequence types belonging to the same serotype. Even though some genetic signatures have been consistently observed in generalists and specialists genomes, the exact mechanisms of host-adaption remain largely unknown.

Epidemiological data collected by the Emilia Romagna Enter-net reference center showed that in swine, the main *Salmonella* reservoir in this region, the second serotype most isolated is Derby, which is only rarely found in humans. The two most prevalent *S*. Derby PFGE profiles in swine, SXB_BS.0204 and SXB_BS.0056, have different prevalence in humans: the proportion of SXB_BS.0204 human isolates among all *S*. Derby human isolates is 10 times lower (0.9%) than that of swine isolates (9.4%), whereas the proportions of SXB_BS.0056 isolates found in humans (13%) and swine (9.7%) are comparable. The percentage of similarity of the two PFGE profiles is above 90%. Thus, isolates from SXB_BS.0204 and SXB_BS.0056 profiles, despite being differentially isolates from humans, are genetically closely related. This context was therefore evaluated as a starting point to investigate host adaptation in *S.* Derby. The limited genetic diversity among isolates from the two PFGE profiles reflects a reduced amount of genetic features needed to be evaluate for their role in host-adaptation.

C. elegans was used to assess if different distributions of SXB_BS.0204 and SXB_BS.0056 in human

are due to differences in virulence detectable in this nematode infection model. SXB_BS.0204 isolates showed the same ability to kill the nematode of SXB_BS.0056 isolates, as well as other isolates with pulsotypes found at the same level in both humans and swine. It has been previously shown that *S*. Typhimurium mutants deleted for genes involved in human pathogenicity exhibit diminished ability to kill the worms. Tested mutants include strains deleted in SPI-1 genes, responsible for invasion of gut epithelial cells (118). However, it was reported that in *C. elegans* no breach of the nematode intestinal epithelium is detected upon infection with *S*. Typhimurium and that intestinal epithelial cells contained only rare intact bacteria and numerous autolysosomes containing bacterial debris (119). Pathogenesis in the nematode of *Salmonella* seems thus not to involve the classical invasive and intracellular lifestyle, but rather the ability to provoke systemic oxidative stress in the host.

Hypothesizing that differences in virulence between SXB_BS.0204 and SXB_BS.0056 isolates could reside in their ability to infect and survive in host cells, invasion and replication assays in human INT-407 and swine IPEC-J2 epithelial cell lines were performed on *S.* Derby isolates with SXB_BS.0204 and SXB_BS.0056 and closely related profiles (CRPs). In human cells, the invasion and replication efficiencies of SXB_BS.0056 isolates were respectively up to 3 and 4 logs higher than those of SXB_BS.0204 isolates. By contrast, in swine IPEC-J2 cells differences in invasion and replication efficiencies between SXB_BS.0204 and SXB_BS.0056 isolates were significantly reduced to just one log. Therefore, SXB_BS.0056, isolated with similar frequency in humans as well as in swine, is more virulent, especially in human cells, than SXB_BS.0204 isolates, which is isolated in humans less frequently than in swine. Isolates with CRPs showed the same virulence phenotype of SXB_BS.0056 isolates (all together called virulent (V) isolates), except for one isolate which infected at the same level of SXB_BS.0204 isolates (non-virulent (NV) isolates).

Differences in host association inferred from the epidemiological data thus correlate with differences in virulence phenotypes observed in *in vitro* cells culture infection assays. These observations, together with the results from *C. elegans* infection assay, prompted us to hypothesize that SXB_BS.0204 isolates are found less frequently in humans than in swine because of their reduced ability to infect human intestinal epithelial cells.

Although Pulsed-field gel electrophoresis is the standard procedure for *Salmonella* typing, its discrimination power is limited to the observation of DNA banding patterns representing only a subset of the genetic differences between strains. To describe the phylogenetic relationship of V and NV isolates with higher resolution, 22 isolates tested in invasion and replication assays and 23 additional isolates with different PFGE profiles were whole genome sequenced, and a SNPs-based phylogenetic tree was generated. The tested isolates cluster separately according to their virulent phenotype, forming the V and NV clades. These results allowed us to speculate that the observed different virulence phenotypes were due to specific genetic determinants shared among isolates from the same clade, and differentiating V and NV isolates. The SXB_BS.0204 isolate N11, which showed higher invasion and replication efficiencies in both cell lines compared to isolates from the same pulsotype, clusters together with NV isolates. Two hypotheses were formulated to explain this scenario. N11 could use the same infection pathway of the other V isolates even if it is phylogenetically closer to NV isolate, or it could infect through a distinct mechanism. Therefore, the isolate was further analyzed both together with the other V isolates and separately.

In order to find the genetic determinants responsible for the observed differences in virulence, the genomes of the 22 isolates tested in invasion and replication assays were analyzed to detect the presence of exclusive genes in V isolates. Host adaptation is often associated with genome decay, including loss of genes related to generalist lifestyle. Our analysis revealed the presence of an

exclusive plasmid in N11 and 69 genes share among the other V isolates, forming SGI-1 and 6 other regions, one of them hypothesized to be located on a plasmid. None of these genes has been previously associated to *Salmonella* pathogenesis, but several hypothetical proteins were annotated. To assess their role in virulence, each region and plasmid was individually deleted and *in vitro* cell infection assays were performed with the generated mutants. We found no difference in invasion and replication efficiencies in human and swine epithelial cells compared to the respective wild type strains, indicating that the phenotype of V isolates is not due to the presence of specific genes.

The ability of *Salmonella* to colonize different hosts has been associated also to the allelic variation of genes involved in pathogenesis. SNPs which discriminate V and NV isolates were thus selected to assess their role in the observed virulence phenotypes. NV isolates carry two non-sense mutations, respectively in *ydiV* and *yhaK* genes. *ydiV* encodes for a repressor of the flagellar genes, required for resistance to host phagocyte oxidase, for suppression of flagellin-mediated killing of macrophages, and for promoting resistance to hydrogen peroxide. *yhaK* encodes for a protein of unknown function, shown to be a good marker for monitoring oxidative stresses in *E. coli*. Both non-sense mutations are located in the middle of the coding sequence, resulting in a truncated product that is likely nonfunctional. As the accumulation of pseudogenes is a recurring feature of host-adapted pathogens, the involvement of these SNPs in virulence was assessed. *ydiV* and *yhaK* allelic variants were individually replaced in ER1175 (V isolate) with the NV alleles. However, these mutants did not exhibit differences in infection of human and swine cells compared to the wild type strain, demonstrating that point mutations in these two genes are not responsible for the observed virulence phenotypes.

Of the 10 missense non conservative mutations that discriminate NV from V isolates, one of them

was found in *hilD*. This gene encodes for the dominant regulator of SPI-1, which encodes for a T3SS that injects effector proteins in the host cells cytosol. In absence of HilD there is no expression of *hilA*, gene that encodes for the direct activator of *invF* and of the T3SS operons. InvF in turn activates genes encoding for the effector proteins translocated in host cells by T3SS. Although SPI-1 independent mechanisms were discovered in the last years, the prevailing paradigm of *Salmonella* pathogenesis asserts that SPI-1 is essential for bacterial invasion of gut epithelial cells, as the effector proteins translocated in host cells by T3SS promote cytoskeletal remodeling and inflammation.

The mutated amino acid in HilD is immediately downstream of one residue predicted to make base-specific contact with DNA. In addition, a strong destabilizing effect was predicted on the protein fold due to the missense mutation. NV isolates were found to carry other two mutations on SPI-1: a missense non-conservative mutation in *sipA*, encoding for an effector protein, and a stop codon in *hilC*, encoding for an additional activator of SPI-1. These two mutations were also found in the isolate N11. As N11 still carries the virulent *hilD* allele and showed higher invasion and replication efficiencies compared to the NV isolates, we hypothesized that mutations in *sipA* and *hilC* are not involved in the virulence phenotype of NV isolates.

The two *hilD* allelic variants were thus analyzed for their role in virulence. The virulent allele of ER1175 (ER1175 WT) was replaced by recombination with the non-virulent allele and vice versa in ER278 (ER278 WT, NV isolate), to generate ER1175::*hilD_nv* and ER278::*hilD_v*, respectively. Both wild type strains were also deleted individually for *hilD* and *invA* (encoding for a main structural component of the T3SS-1) to assess respectively if the deletion of the entire gene and the loss of a functional T3SS-1 confer the same virulence phenotype of the *hilD* non-virulent allele. Results showed two different scenarios in human and swine cells. The nsSNP in *hilD* is the cause of different invasion and replication efficiencies in human cells. Carrying *hilD_nv* allele elicits the

same virulence phenotype of *hilD* deletion and the absence of a functional T3SS-1 in ER1175. ER278 WT showed only slightly higher invasion and replication efficiencies compared to the their *hilD* and *invA* deleted mutants, whereas ER278::*hilD_v* ability to infect is even higher than that of ER1175 WT. These results suggested that *hilD_nv* encodes for a nonfunctional protein.

On the other hand, swine cells infection assays showed that in ER1175 the introduction of *hilD_nv* as well as the *hilD* deletion does not cause changes in the virulence phenotype, whereas the *invA* deletion triggers a decrease in virulence. These results suggest that V isolates might be able to induce T3SS expression independently from HilD, suggesting an alternative method to activate SPI-1 during infection of swine cells. Alternatively, the V isolates might be able to infect swine cells by a different mechanism. No differences were found among ER278 WT, ER278Δ*hilD* and ER278Δ*invA*, but *hilD_v* insertion causes an increase in both invasion and replication compared to ER278 WT. It may be possible that NV isolates are adapted to not use SPI-1, even if it remains still able to be activated by the *hilD_v* encoded regulator.

Overall, these results indicate that the two *hilD* allelic variants explain virulence differences between V and NV isolates in human cells, but they do not fully explain the differences observed in swine cells.

The virulence phenotypes of mutants deleted for *hilD* and *invA* observed in human cells suggest that *hilD_nv* encodes for a nonfunctional protein which does not activate the expression of SPI-1 genes. To verify this hypothesis, the expression profiles of ER1175 WT, ER278 WT, ER1175::*hilD_nv*, ER278::*hilD_v*, ER1175Δ*hilD*, ER278Δ*hilD* strains were compared. No differences in the whole transcriptome were observed between both ER1175 and ER278 deleted for *hilD* or carrying *hilD_nv*. By contrast, in both ER1175 and ER278 carrying *hilD_v* compared to the respective mutants deleted for *hilD* or carrying *hilD_nv*, SPI-1 genes are upregulated, as well as

SPI-5 encoded *sopB* and *pipC*, SPI-4 encoded *siiA*-*siiD* genes, and 3 CDS encoding for hypothetical proteins. *sopB* and *pipC* genes encode respectively for an effector protein, responsible for actin rearrangements, and its specific chaperone. The two genes are regulated by *invF* and *sicA*, which are activated by *hilA*, in turn dependent on *hilD* activation. SPI-4 encodes for a type 1 secretion system (T1SS) and an associated giant non-fimbrial adhesine which mediates attachment to epithelial cells. *siiA*, the first gene in SPI-4, was proposed to be activated by HilD, as well as by the HilD regulated HilC, RtsA, HilA, and SprB (138).

SPI-1, SPI-4 and SPI-5 genes were found to be downregulated also in *S*. Typhimurium Δ*hilD* compared to the wild type (103). By contrast the 3 CDS encoding for hypothetical proteins, even if present also in the *S*. Typhimurium genome, were not found differentially expressed in the two *S*. Typhimurium strains. In addition, *hilD* deletion in *S*. Typhimurium causes downregulation of SPI-2 encoded and associated genes, putative virulence factors and T3SS-independent virulence factors compared to the wild type. These genes were not found differentially expressed in ER278 and ER1175 carrying *hilD_v* compared to both *hilD* deleted and *hilD_nv* carrying strains. These results allow to hypothesize that the set of genes regulated by *hilD* in *S*. Derby might not totally match with that of *S*. Typhimurium.

The introduction of *hilD_v* in ER278 triggers changes in expression of about 250 genes other than those shared with ER1175 WT. NV isolates normally express a non-functional HilD, therefore it is possible that the expression of the *hilD_v* encoded regulator, other than activating the expression of *hilD* dependent genes, could affects the stability of the cell transcription process. An in-depth examination of these results will define the effect of *hilD_v* expression in NV isolates, potentially pointing out also new genes subject to HilD regulation.

To verify if SPI-1 is downregulated in presence of *hilD_nv* even during human cells infection, bacterial RNA of ER1175 WT, ER1175::*hilD_nv* and ER1175Δ*hilD* was extracted at three different

infection times and expression analysis was performed on *hilD*, other 6 SPI-1 genes and *pagN*. The expression of *pagN*, known to support adhesion and invasion of epithelial cells, is regulated by the PhoPQ system, so it was added to the analysis as indicator of the expression of virulence genes regulated independently from HilD. During human cells infection ER1175 WT increases expression of *hilD* and SPI-1 genes, while the expression of these genes in ER1175::*hilD_nv* and ER1175Δ*hilD* remains constantly low overtime. Therefore *hilD_nv* allele encodes for a product unable to activate SPI-1 genes expression, both in *in vitro* conditions and during human cells infection. The dynamic of *pagN* expression is different from those of SPI-1 genes: its expression increases overtime even in ER1175::*hilD_nv* and ER1175Δ*hilD*, but in a lower extent than in ER1175 WT, suggesting that HilD could be involved in the amplification of *pagN* expression. Intriguingly, no differences in *pagN* expression were detected in the whole transcriptome analysis made on RNA extracted during early stationary phase. These data suggest that the *in vitro* conditions used to extract RNA for whole transcriptome analysis do not fully recapitulate the conditions encountered by *Salmonella* during human cells infection.

We showed that the presence of *hilD_nv* causes downregulation of SPI-1 genes expression, responsible for the lower ability of NV isolates to infect human cells. Do *hilD* allelic variants trigger different virulence phenotypes even *in vivo*? The most widely used animal model for studying *Salmonella* pathogenesis is a susceptible mouse which develops both systemic infection, due to a mutation in Nramp1 protein which is correlated to immune responses, and intestinal inflammation, because of a pretreatment with streptomycin, which reduces colonization by the commensal intestinal microbes responsible for intrinsic resistance of mice to *Salmonella*. It is well established that SPI-1 plays a key role in eliciting intestinal inflammation in the mouse model of colitis during *S*. Typhimurium infection. A mutant deleted for *invG*, encoding for a structural component of T3SS-1, as well as mutants deleted individually for *sipA*, *sopE* and *sopE2*, encoding

for T3SS-1 translocated effectors, causes only mild inflammation (137). SPI-1 is also involved in systemic virulence, as it has been reported that the bacterial loads in spleen and liver of *S*. Typhimurium SPI-1 mutants are lower than those of the wild type strains (29, 139, 140). The proposed mechanism is that SPI-1 mutants have reduced abilities to traverse the intestinal epithelium and thus to reach systemic organs. Additionally, SPI-1 mutants cause only mild gut inflammation and thus are less able to cause phagocytes transmigration into the intestinal lumen, thus reducing chances to disseminate within this immune cells.

These findings are corroborated by the observation that SPI-1 mutants, when administered intraperitoneally to mice, show no differences in their ability to colonize the spleen compared to the wild type strain (139). *Salmonella* follows a short course of infection after intraperitoneal infection, trafficking within mononuclear cells through the lymph nodes in the peritoneal cavity to the blood stream and then spleen and liver, thus bypassing the intestinal route which involves SPI-1-driven host cells invasion.

To assess if *hilD_nv* causes altered *S*. Derby pathogenicity compared to *hilD_v* in the mouse model, streptomycin-pretreated *Nramp1*-deficient mice were infected with ER1175 WT, ER1175::*hilD_nv*, ER1175Δ*hilD,* and ER278 WT. No differences among strains were found in the cecal bacterial loads as well as in the expression of the pro-inflammatory genes *Cxcl1* and *Nos2*. Moreover, *Cxcl1* and *Nos2* were not upregulated in mice infected with any of the tested strains, compared to uninfected mice. These results suggest that all tested strains are able to replicate in the cecum, but without causing inflammation. There are at least two alternative explanations for the phenotype observed. On the one hand, it was already observed that a common feature of host adaptation is the loss of virulence for secondary hosts. For example, isolates belonging to the pigeon-restricted *S*. Typhimurium definitive type 2 (DT2) showed attenuated virulence in mice compared to isolates from sequence type 19, considered the ancestral genotype (141). On the

other hand, it was observed that *Salmonella* serovars adapted or restricted to a specific host have evolved from causing enterocolitis, relied to intestinal inflammation, to an alternative lifestyle that involves blunting of the inflammatory response and dissemination to systemic sites. The host restricted serotypes Typhi and Paratyphi A, Pullorum and Gallinarum cause systemic typhoid diseases respectively in human and poultry. Suar et al (50) showed, using the same mice model of *Salmonella* infection used in this work, that *S*. Enteritidis and *S*. Typhimurium, host unrestricted serotypes, cause pronounced colitis, whereas only mild colitis is observed upon infection with different strains of serovar Pullorum. Moreover, *S*. Paratyphi strains showed no ability to cause intestinal inflammation, and variability in causing colitis was observed among S. Dublin and S. Gallinarum tested strains. The absence of intestinal inflammation upon *S*. Derby infection, regardless of the *hilD* allelic variant or even the gene deletion, may thus indicate that this serovar has evolved to lose ability to infect a secondary host and/or to adopt a mechanism of pathogenesis similar to the other host adapted serovars.

Even in *C. elegans* infection assays the *hilD* allelic variants do not cause changes in virulence, as *S*. Derby V and NV isolates kill the nematode at the same extent. By contrast, a *S*. Typhimurium mutant deleted for *hilD* exhibits reduces virulence towards the nematode compared to the respective wild type (118). These data allow to speculate that the *S*. Derby virulence in *C. elegans*, in contrast to that of *S*. Typhimurium, could be independent of *hilD*. It should be considered that the invasion of epithelial cells seems not to be crucial for the pathogenesis in the nematode, therefore the role of *hilD* in virulence remains unclear in this model.

We found that ER1175 WT load in spleen and liver is lower than that of ER1175Δ*hilD*, even if this difference was significant only in the spleen. HilD, and in turn SPI-1 seem thus to be not necessary, and even detrimental, for *S*. Derby dissemination. This result is in contrast with what is seen in mice infected with SPI-1-deficient *S*. Typhimurium, where the extent of dissemination is directly

correlated with the ability to cause gut colonization and inflammation. The reason may reside in the dual behavior of *Salmonella* towards macrophages: *Salmonella* needs to reside within macrophages for systemic dissemination, but it also induces macrophage death to contribute to inflammatory milieu and to replicate in the gastrointestinal lumen. As one mechanism of induction of cell death is pyroptosis, which can be induced by secretion of flagellin and SipB into host cell cytosol via T3SS-1 (142, 143), the absence of SPI-1 expression could potentially lead to reduced levels of pyroptosis, which would then result into higher numbers of infected macrophages that could migrate to the liver and spleen.

ER1175::*hilD_nv* accumulates in spleen and liver at an intermediate level between ER1175 WT and ER1175Δ*hilD*, but no significant differences between ER1175::*hilD_nv* and both strains were detected. ER278 WT and ER1175 WT disseminate similarly, despite the fact that they carry distinct *hilD* allelic variants. In mouse infection assays, the effect of carrying *hilD_nv* is thus not well-defined. In evaluating these results, it should be considered that the mouse model of infection has been developed on *S.* Typhimurium. This serovar, as already illustrated, shows higher virulence than host-adapted and host-restricted serotypes in mice, suggesting that host adaptation causes loss of virulence in secondary hosts. For this reason, it can be speculated that the mouse infection model does not allow to fully investigate the pathogenicity of *S*. Derby.

Overall, this work emphasizes that allelic variants in virulence genes can cause altered virulence phenotypes that potentially relate to host specificity. The already reported mechanisms of host-adaptation associated to nsSNPs are related to changes in surface or effector proteins involved in recognition and invasion of host cells (55,56). Here, we found that a point mutation in the virulence genes regulator *hilD* might be involved in the low prevalence in humans of the

SXB_BS.0204 PFGE profile studied. The differences in virulence phenotypes observed in human cell line infection assays between isolates carrying distinct *hilD* allelic variants suggest that changes in host association may occur by point mutations that, affecting the functionality of a single gene, have an impact on the expression of an extended virulence gene repertoire.

We observed that *hilD* is not essential for *S*. Derby infection of swine cells. This is in line with the observation that SXB_BS.0204 and SXB_BS.0056 isolates were found in the same proportion in swine, even though SXB_BS.0204 isolates carry *hilD_nv* which encodes for a nonfunctional activator of SPI-1. The *hilD* point mutation of NV isolates is fixed in the population, as it is found in several sporadic isolates spread in time and space. Therefore, the classical genetic pathway considered essential for invasion does not explain the infection of swine cells. Two hypotheses were formulated to explain this result: it is possible that *S*. Derby invades swine cells by adopting an *hilD* independent mechanism to activate SPI-1, or an alternative mechanism that does not involve SPI-1. Further experiments are needed to fully elucidate the mechanism of swine epithelial cell invasion.

The expression analysis indicated that the HilD regulon might be not completely described, and that distinct genes could be affected by HilD regulation in *S*. Derby compared to *S*. Typhimurium. Additional analysis needs to be performed to confirm and characterize the observed differences that might be related to different host association and pathogenesis among the two serovars.

Future works will also focus on investigating the virulence phenotypes observed in mice, and in particular why *S*. Derby seems to not cause gut inflammation and why the deletion of *hilD* causes a different phenotype in *S*. Derby compared to *S*. Typhimurium.

Together, this study has provided new information about mechanisms of host adaptation by *Salmonella*.

5. Bibliography

1. Schultz M. 2010. Theobald Smith. Emerging Infectious Diseases, 14:1940–1942.

2. Grimont PAD, Weill FX. 2007. Antigenic formulae of the *Salmonella* serovars, 9th revision. World Health Organization Collaborating Center for Reference and Research on *Salmonella*, Pasteur Institute, Paris, France.

3. Fàbrega A, Vila J. 2013. *Salmonella enterica* Serovar Typhimurium Skills To Succeed in the Host: Virulence and Regulation. Clinical Microbiology Reviews, 26:308–341.

4. Yan SS, Pendrak ML, Abela-Ridder B, Punderson JW, Fedorko DP, et al. 2003. An overview of *Salmonella* typing Public health perspectives. Clinical and Applied Immunology Reviews, 4:189–204.

5. Deng X, den Bakker HC, Hendriksen RS. 2016. Genomic Epidemiology: Whole-Genome-Sequencing–Powered Surveillance and Outbreak Investigation of Foodborne Bacterial Pathogens. Annual Review of Food Science and Technology, 7:353–74.

6. Schürch AC, Arredondo-Alonso S, Willems RJL, Goering RV. 2018. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. Clinical Microbiology and Infection, 24:350-354.

7. Seemann T. 2015. Snippy: fast bacterial variant calling from NGS reads. https://github.com/tseemann/snippy.

8. Davis S, Pettengill JB, Luo Y, Payne J, Shpuntoff A, Rand H, Strain E. 2015. CFSAN SNP Pipeline: an automated method for constructing SNP matrices from next-generation sequence data. PeerJ Computer Science, 1:e20.

9. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution, 30(12):2725–29.

10. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics, 22(21):2688–90.

11. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2: approximately maximum-likelihood trees for large alignments. PLOS ONE, 5(3):e9490.

12. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution, 29(8):1969–73.

13. Foster JW, Hall HK. 1991. Inducible pH homeostasis and the acid tolerance response of *Salmonella* typhimurium. Journal of Bacteriology, 173:5129–5135.

14. LaRock DL, Chaudhary A, Miller SI. 2015. *Salmonella*e interactions with host processes. Nature

Reviews Microbiology, 13:191-205.

15. Jones GW, Richardson LA, Uhlman D. 1981. The invasion of HeLa cells by *Salmonella typhimurium*: reversible and irreversible bacterial attachment and the role of bacterial motility. Journal of general microbiology, 127:351–360.

16. Das C, Mokashi C, Mande SS, Saini S. 2018. Dynamics and Control of Flagella Assembly in *Salmonella* typhimurium. Frontiers in Cellular and Infection Microbiology, 8:36.

17. Baumler AJ, Tsolis RM, Heffron F. 1997. Fimbrial adhesins of *Salmonella typhimurium*. Role in bacterial interactions with epithelial cells. Advances in Experimental Medicine and Biology, 412:149–158.

18. Romling U, Bian Z, Hammar M, Sierralta WD, Normark S. 1998. Curli fibers are highly conserved between *Salmonella typhimurium* and *Escherichia coli* with respect to operon structure and regulation. Journal of Bacteriology, 180:722–731.

19. Collinson SK, Doig PC, Doran JL, Clouthier S, Trust TJ, et al. 1993. Thin, aggregative fimbriae mediate binding of *Salmonella enteritidis* to fibronectin. Journal of Bacteriology, 175:12–18.

20. Chessa D, Dorsey CW, Winter M, Baumler AJ. 2008. Binding specificity of *Salmonella* plasmid-encoded fimbriae assessed by glycomics. Journal of Biological Chemistry. 283:8118–8124.

21. Chessa D, Winter MG, Jakomin M, Baumler AJ. 2009. *Salmonella enterica* serotype Typhimurium Std fimbriae bind terminal alpha(1,2)fucose residues in the cecal mucosa. Molecular Microbiology, 71:864–875.

22. Kiss T, Morgan E, Nagy G. 2007. Contribution of SPI-4 genes to the virulence of *Salmonella enterica*. FEMS Microbiology Letters, 275:153–159.

23. Latasa C, Roux A, Toledo-Arana A, Ghigo JM, Gamazo C, et al. 2005. BapA, a large secreted protein required for biofilm formation and host colonization of *Salmonella enterica* serovar Enteritidis. Molecular Microbiology, 58:1322–1339.

24. Dorsey CW, Laarakker MC, Humphries AD, Weening EH, Baumler AJ. 2005. *Salmonella enterica* serotype Typhimurium MisL is an intestinal colonization factor that binds fibronectin. Molecular Microbiology, 57:196–211.

25. Lara-Tejero M, Galan JE. 2009. *Salmonella enterica* serovar Typhimurium pathogenicity island 1-encoded type III secretion system translocases mediate intimate attachment to nonphagocytic cells. Infection and Immunity, 77:2635–2642.

26. Zhou D, Galán J. 2001. *Salmonella* entry into host cells: the work in concert of type III secreted effector proteins. Microbes and Infection, 3:1293-1298.

27. Hurley D, McCusker MP, Fanning S, Martins M. 2014. *Salmonella*–host interactions - modulation of the host innate immune system. Frontiers in Immunology, 5:481.

28. Golubeva YA, Sadik YA, Ellermeier JR, Slauch JM. 2012. Integrating Global Regulatory Input Into the *Salmonella* Pathogenicity Island 1 Type III Secretion System. Genetics, 190:79-90.

29. Ellermeier CD, Ellermeier JR, Slauch JM. 2005. HilD, HilC and RtsA constitute a feed forward loop that controls expression of the SPI1 type three secretion system regulator hilA in *Salmonella* enterica serovar Typhimurium. Molecular Microbiology, 57: 691–705.

30. Schechter LM, Damrauer SM, Lee CA. 1999. Two AraC/XylS family members can independently counteract the effect of repressing sequences upstream of the hilA promoter. Molecular Microbiology, 32: 629–642.

31. Saini S, Ellermeier JR, Slauch JM, Rao CV. 2010. The Role of Coupled Positive Feedback in the Expression of the SPI1 Type Three Secretion System in *Salmonella*. PLoS Pathogens, 6(7):e1001025.

32. Hautefort I, Thompson A, Eriksson Ygberg S, Parker ML, Lucchini S, at al. 2008. During infection of epithelial cells *Salmonella enterica* serovar Typhimurium undergoes a time-dependent transcriptional adaptation that results in simultaneous expression of three type 3 secretion systems. Cellular Microbiology, 10:958–984.

33. Bajaj V, Lucas RL, Hwang C, Lee CA. 1996. Co-ordinate regulation of *Salmonella typhimurium* invasion genes by environmental and regulatory factors is mediated by control of *hilA* expression. Molecular Microbiology, 22:703–714.

34. Kroger C, Colgan A, Srikumar S, Handler K, Sivasankaran SK, et al. 2013. An infection-relevant transcriptomic compendium for *Salmonella* enterica Serovar Typhimurium. Cell Host & Microbe; 14:683-95.

35. Sánchez-Romero MA, Casadesús J. 2018. Contribution of SPI-1 bistability to *Salmonella* enterica cooperative virulence: insights from single cell analysis. Scientific Reports**,** 8:14875.

36. Rivera-Chávez F, Lopez CA, Zhang LF, García-Pastor L, Chávez-Arroyo A, et al. 2016. Energy taxis toward host-derived nitrate supports a *Salmonella* pathogenicity island 1-independent mechanism of invasion. mBio, 7(4):00960-16.

37. Zhang S, Santos RL, Tsolis RM, Stender S, Hardt WD, et al. 2002. The *Salmonella enterica* serotype Typhimurium effector proteins SipA, SopA, SopB, SopD, and SopE2 act in concert to induce diarrhea in calves. Infection and Immunity, 70:3843–3855.

38. Fu Y, Galan JE. 1999. A *Salmonella* protein antagonizes Rac-1 and Cdc42 to mediate host-cell

recovery after bacterial invasion. Nature 401: 293–297.

39. Spees AM, Lopez CA, Kingsbury DD, Winter SE, Bäumler AJ. 2013. Colonization resistance: battle of the bugs or Ménage à Trois with the host? PLoS Pathogens, 9:e1003730.

40. Kuhle V, Hensel M. 2004. Cellular microbiology of intracellular *Salmonella enterica*: functions of the type III secretion system encoded by *Salmonella* pathogenicity island 2. Cellular and Molecular Life Sciences, 61:2812–2826.

41. Hernandez LD, Hueffer K, Wenk MR, Galan JE. 2004. *Salmonella* modulates vesicular traffic by altering phosphoinositide metabolism. Science, 304:1805–1807.

42. Jackson LK, Nawabi P, Hentea C, Roark EA, Haldar K. 2008. The *Salmonella* virulence protein SifA is a G protein antagonist. PNA S, 105:14141–14146.

43. Brawn LC, Hayward RD, Koronakis V. 2007. *Salmonella* SPI1 effector SipA persists after entry and cooperates with a SPI2 effector to regulate phagosome maturation and intracellular replication. Cell Host Microbe, 1:63–75.

44. Guignot J, Caron E, Beuzon C, Bucci C, Kagan J, et al. 2004. Microtubule motors control membrane dynamics of *Salmonella* containing vacuoles. Journal of Cell Science, 117:1033–1045.

45. Helaine S, Cheverton AM, Watson KG, Faure LM, Matthews SA, et al. 2014. Internalization of *Salmonella* by macrophages induces formation of nonreplicating persisters. Science, 10:204-8

46. Malik-Kale P, Winfree S, Steele-Mortimer O. 2012. The bimodal lifestyle of intracellular *Salmonella* in epithelial cells: replication in the cytosol obscures defects in vacuolar replication. PLoS ONE, 7(6):e38732.

47. Fink SL, Cookson BT. 2007. Pyroptosis and host cell death responses during *Salmonella* infection. Cellular Microbiology, 9:2562–2570.

48. Behnsen J, Perez-Lopez A, Nuccio S, Raffatellu M. 2015. Exploiting host immunity: the *Salmonella* paradigm. Trends in Immunology, 36:112-120.

49. Hersh D, Monack DM, Smith MR, Ghori N, Falkow S, et al. 1999. The Salmonella invasion SipB induces macrophage apoptosis by binding to caspase-1. PNAS, 96:2396–2401.

50. Suar M, Jantsch J, Hapfelmeier S, Kremer M, Stallmach T, et al. 2006. Virulence of broad- and narrowhost-range *Salmonella* enterica serovars in the streptomycin-pretreated mouse model. Infection and Immunology, 74: 632-644.

51. Bäumler AJ, Fang FC. 2013. Host Specificity of Bacterial Pathogens. Cold Spring Harbour Perspective in Medicine, 3:a010041.

52. Tanner JR, Kingsley RA. 2018. Evolution of *Salmonella* within hosts. Trends in Microbiology,

1591:13

53. Branchu P, Matt Bawn M, Kingsley RA. 2018. Genome variation and molecular epidemiology of *Salmonella enterica* serovar Typhimurium Pathovariants. Infection and Immunology, 6:e00079-18.

54. Langridge GC, Fookes M, Connor TR, Feltwell T, Feasey N, et al. 2015. Patterns of genome evolution that have accompanied host adaptation in *Salmonella*. PNAS, 112:863–868.

55. Eswarappa SM, Janice J, Nagarajan AG, Balasundaram SV, Karnam G, et al. 2008. Differentially evolved genes of *Salmonella* Pathogenicity Islands: Insights into the mechanism of host specificity in *Salmonella*. PloS One, 3:e3829.

56. Yue M, Han X, De Masi L, Zhu C, Ma X, et al. 2015. Allelic variation contributes to bacterial host specificity. Nature Communications, 6:8754.

57. Feasey NA, Dougan G, Kingsley RA, Heyderman RS, Gordon MA. 2012. Invasive non-typhoidal *Salmonella* disease: an emerging and neglected tropical disease in Africa. Lancet, 379(9835):2489–2499.

58. Lawley TD, Bouley DM, Hoy YE, Gerke C, Relman DA, Monack DM. 2008. Host transmission of *Salmonella* enterica serovar Typhimurium is controlled by virulence factors and indigenous intestinal microbiota. Infection and Immunity, 76: 403–416.

59. Kothapalli S, Nair S, Alokam S, Pang T, Khakhria R,Woodward D, Johnson W, Stocker BA, Sanderson KE, Liu SL. 2005. Diversity of genome structure in *Salmonella* enterica serovar Typhi populations. Journal of Bacteriology, 187: 2638–2650.

60. Liu WQ, Liu GR, Li JQ, Xu GM, Qi D, et al. 2007. Diverse genome structures of *Salmonella* Paratyphi C. BMC Genomics, 8: 290.

61. Wu KY, Liu GR, Liu WQ, Wang AQ, Zhan S, et al. 2005. The genome of *Salmonella* enterica serovar Gallinarum: Distinct insertions/deletions and rare rearrangements. Journal of Bacteriology, 187: 4720–4727.

62. Wangdi T, Lee C-Y, Spees AM, Yu C, Kingsbury DD, et al. (2014) The Vi Capsular Polysaccharide Enables *Salmonella enterica* Serovar Typhi to Evade Microbe-Guided Neutrophil Chemotaxis. PLoS Pathogens, 10(8): e1004306.

63. Pickard D, Wain J, Baker S, Line A, Chohan S, et al. 2003. Composition, acquisition, and distribution of the Vi exopolysaccharide-encoding *Salmonella* enterica pathogenicity island SPI-7. Journal of Bacteriology, 185: 5055–5065.

64. Hibbing ME, Fuqua C, Parsek MR, Peterson SB. 2010. Bacterial competition: surviving and thriving in the microbial jungle. Nature Reviews Microbiology, 8:15–25.

65. Kingsley RA, Bäumler AJ. Host adaptation and the emergence of infectious disease: the *Salmonella* paradigm. Molecular Microbiology, 36(5):1006-1014.

66. Kirchner JW, Roy BA. 2002. Evolutionary implications of host-pathogen specificity: Fitness consequences of pathogen virulence traits. Evolutionary ecology research, 4:27–48.

67. EFSA. 2017. The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2016

68. Hauser E, Hebner F, Tietze E, Helmuth R, Junker E, et al. 2011. Diversity of *Salmonella* enterica serovar Derby isolated from pig, pork and humans in Germany. International Journal of Food Microbiology, 151:141–149.

69. Wang B, Wang C, McKean JD, Logue CM, Gebreyes WA, et al. 2011. *Salmonella* enterica in swine production: assessing the association between amplified fragment length polymorphism and epidemiological units of concern. Applied and Environmental Microbiology, 77:8080–8087.

70. Cai Y, Tao J, Jiao Y, Fei X, Zhou L, et al. 2016. Phenotypic characteristics and genotypic correlation between *Salmonella* isolates from a slaughterhouse and retail markets in Yangzhou, China. International Journal of Food Microbiology, 222:56–64.

71. Hayward MR, Petrovska L, Jansen VA, Woodward MJ. 2016. Population structure and associated phenotypes of *Salmonella* enterica serovars Derby and Mbandaka overlap with host range. BMC Microbiology, 16:15.

72. Zheng H, Hu Y, Li Q, Tao J, Cai Y, et al. 2017. Subtyping *Salmonella* enterica serovar Derby with multilocus sequence typing (MLST) and clustered regularly interspaced short palindromic repeats (CRISPRs). Food Control 73(Part B), 474–484.

73. Sévellec Y, Vignaud M-L, Granier SA, Lailler R, Feurer C, et al. 2018.Polyphyletic Nature of *Salmonella* enterica Serotype Derby and Lineage-Specific Host-Association Revealed by Genome-Wide Analysis. Frontiers in Microbiology, 9:891.

74. Amavisit P, Lightfoot D, Browning GF, Markham PF. 2003. Variation between pathogenic serovars within *Salmonella* pathogenicity islands. Journal of Bacteriology, 185:3624–3635.

75. Hayward MR, AbuOun M, La Ragione RM, Tchorzewska MA, Cooley WA, et al. 2014. SPI-23 of S. Derby: role in adherence and invasion of porcine tissues. PLoS One, 9:e107857.

76. Standard Operating Procedure for PulseNet PFGE of Escherichia coli O157:H7, Escherichia coli non-O157 (STEC), Salmonella serotypes, Shigella sonnei and Shigella flexneri 2013. Available from: https://www.cdc.gov/pulsenet/PDF/ecoli-shigella-salmonella-pfge-protocol-508c.pdf.

77. Maloy, SR, Stewart VJ, Taylor RK. 1996. Genetic Analysis of Pathogenic Bacteria: a Laboratory

Manual. Cold Spring Harbor Laboratory Press, Plainview, NY.

78. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. PNA S, 97:6640–6645.

79. Stiernagle T. 2006. Maintenance of C. elegans. WormBook, ed. The C. elegans Research Community, WormBook.

80. Kaplan EL, Meier P. 1958. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 53:457–481.

81. Harrington D. 2005. Linear Rank Tests in Survival Analysis. Encyclopedia of Biostatistics, 5.

82. Vaida, Florin and Fitzgerald, Anthony and DeGruttola, Victor. 2007. Efficient Hybrid EM for nonlinear mixed effects models with censored response. Computational Statistics and Data Analysis, 51:5718-5730.

83. European Committee on Antimicrobial Susceptibility Testing. 2016. Breakpoint tables for interpretation of MICs and zone diameters. Version 6.0. European Committee on Antimicrobial Susceptibility Testing, Växjö, Finland. http://www.eucast.org/clinical_breakpoints/.

84. Zhang JH, Chung TD, Oldenburg KR. 1999. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. Journal for Biomolecular Screening, 4:67–73.

85. Carnevali C, Morganti M, Scaltriti E, Bolzoni L, Pongolini S, Casadei G. 2016. Occurrence of mcr-1 in colistin-resistant *Salmonella* enterica isolates recovered from humans and animals in Italy, 2012 to 2015. Antimicrobial Agents and Chemotherapy, 60:7532–7534

86. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

87. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, et al. 2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. Journal of Computational Biology 20(10):714-737.

88. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, et al. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics, 11(1):119.

89. Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Research, 13:2178–2189.

90. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, et al. 2008. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics, 9:75.

91. Clausen PTLC, Aarestrup FM, Lund O. 2018. Rapid and precise alignment of raw reads against redundant databases with KMA. BMC Bioinformatics, 19:307.

92. Arndt D, Grant J, Marcu A, Sajed T, Pon A, et al. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Research, 44(W1):W16-21.

93. Seemann T. 2015. Snippy: fast bacterial variant calling from NGS reads. https://github.com/tseemann/snippy.

94. Heng L. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, 1303.3997.

95. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. 2009. The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics, 25(16):2078-9.

96. Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv, 1207.3907.

97. Darling AC, Mau B, Blattner FR, and Perna NT. 2014. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Research, 14:1394-1403.

98. Stamatakis A. 2014. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30(9):1312-3.

99. Karunakaran R, Mauchline TH, Hosie AH, Poole PS. 2005. A family of promoter probe vectors incorporating autofluorescent and chromogenic reporter proteins for studying gene expression in Gram-negative bacteria. Microbiology, 151:3249-56.

100. Lopez-Garrido J, Puerta-Fernandez E, Casadesus J. A eukaryotic-like 3 untranslated region in Salmonella enterica hilD mRNA. Nucleic Acids Research, 42(9):5894–5906.

101. You FM, Huo N, Gu YQ, Luo M, Ma Y, et al. 2008. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinformatics, 9:253.

102. Wangkumhang P, Chaichoompu K, Ngamphiw C, Ruangrit U, Chanprasert J, et al. 2007. WASP: a Web-based Allele-Specific PCR assay designing tool for detecting SNPs and mutations. BMC Genomics, 8:275-283.

103. Colgan AM, Kröger C, Diard M, Hardt W-D, Puente JL, et al. (2016) The Impact of 18 Ancestral and Horizontally-Acquired Regulatory Proteins upon the Transcriptome and
sRNA Landscape of Salmonella enterica serovar Typhimurium. PLoS Genetics 12(8): e1006258.

104. ENCODE: Encyclopedia of DNA Elements. 2017. http://www.encodeproject.org.

105. Krueger F. 2010. Trim Galore! is available online at: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

106. Li H. 2016. Minimap and Miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics, 32(14):2103-2110.

107. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 26(6):841–842.

108. Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. Nature Methods, 12:357–60.

109. Anders S, Pyl PT, Huber W. 2014. HTSeq — A Python framework to work with high-throughput sequencing data. Bioinformatics, 31(2):166–169.

110. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12):550.

111. Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Research, 29:e45.

112. Pfaffl MW, Horgan GW, Dempfle L. 2002. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Research, 30:e36.

113. Bellamy R. 1999. The natural resistance–associated macrophage protein and susceptibility to intracellular pathogens. Microbes and Infection, 1:23-7.

114. Barthel M, Hapfelmeier S, Quintanilla-Martínez L, Kremer M, Rohde M, et al. 2003. Pretreatment of mice with streptomycin provides a *Salmonella enterica* serovar Typhimurium colitis model that allows analysis of both pathogen and host. Infection and Immunity, 71(5):2839–58.

115. Thiennimitr P, Winter SE, Winter MG, Xavier MN, Tolstikov V, et al. 2011. Intestinal inflammation allows *Salmonella* to use ethanolamine to compete with the microbiota. PNAS, 108(42):17480–17485.

116. Aballay A, Yorgey P, Ausubel FM. 2000. *Salmonella* Typhimurium proliferates and establishes a persistent infection in the intestine of *Caenorhabditis elegans*. Current Biology, 10:1539–1542.

117. Labrousse A, Chauvet S, Couillault C, Kurz CL, Ewbank JJ. 2000. *Caenorhabditis elegans* is a model host for *Salmonella* Typhimurium. Current Biology, 10:1543–1545.

118. Tenor JL, McCormick BA, Ausubel FM, Aballay A. 2004. *Caenorhabditis elegans*-based screen identifies *Salmonella* virulence factors Required for conserved host-pathogen interactions. Current Biology, 14:1018–1024.

119. Sem X, Rhen M. 2012. Pathogenicity of *Salmonella enterica* in *Caenorhabditis elegans* relies on disseminated oxidative stress in the infected host. PLoS ONE, 7(9):e45417.

120. Doublet B, Boyd D, Mulvey MR, Cloeckaert A. 2005. The *Salmonella* genomic island 1 is an

integrative mobilizable element.

121. Wada T, Morizane T, Abo T, Tominaga A, Inoue-Tanaka K, Kutsukake K. 2011. EAL domain protein YdiV acts as an anti-FlhD4C2 factor responsible for nutritional control of the flagellar regulon in *Salmonella enterica* Serovar Typhimurium. Journal of Bacteriology, 193:1600-1611.

122. Hisert KB, MacCoss M, Shiloh MU, Darwin KH, Singh S. 2005. A glutamate-alanine-leucine (EAL) domain protein of *Salmonella* controls bacterial survival in mice, antioxidant defence and killing of macrophages: role of cyclic diGMP. Molecular Microbiology, 56:1234-1245.

123. Gurmu D, Lu J, Johnson KA, Nordlund P, Holmgren A, Erlandsen H. 2009. The crystal structure of the protein YhaK from *Escherichia coli* reveals a new subclass of redox sensitive enterobacterial bicupins. Proteins: Structure, Function, and Bioinformatics, 74:18–31.

124. Zhou D, Galán J. 2001. *Salmonella* entry into host cells: the work in concert of type III secreted effector proteins. Microbes and Infection, 3:1293-1298.

125. Martinez LC, Banda MM, Fernandez-Mora M, Santana FJ, Bustamante VH. 2014. HilD induces expression of *Salmonella* pathogenicity island 2 genes by displacing the global negative regulator H-NS from ssrAB. Journal of bacteriology, 196:3746-3755

126. Main-Hester KL, Colpitts KM, Thomas GA, Fang FC, Libby SJ. 2008. Coordinate regulation of Salmonella pathogenicity island 1 (SPI1) and SPI4 in Salmonella enterica serovar Typhimurium. Infection and Immunity, 76:1024-1035.

127. Singer HM, Kühne C, Deditius JA, Hughes KT, Erhardt M.2014. The Salmonella Spi1 Virulence Regulatory Protein HilD Directly Activates Transcription of the Flagellar Master Operon flhDC. Journal of Bacteriology 196(7):1448-1457.

128. Eichelberg K, Hardt WD, Galan JE. 1999. Characterization of SprA, an AraC-like transcriptional regulator encoded within the *Salmonella* typhimurium pathogenicity island 1. Molecular Microbioly, 33(1):139-152.

129. Petrone BL, Stringer AM, Wade JT. 2013. Identification of HilD-Regulated Genes in Salmonella enterica Serovar Typhimurium. Journal of Bacteriology, 196(5):1094 –1101.

130. Schechter LM, Damrauer SM, Lee CA. 1999. Two AraC/XylS family members can independently counteract the effect of repressing sequences upstream of the hilA promoter. Molecular Microbiology, 32: 629–642.

131. Schechter LM, Lee CA. 2001. AraC/XylS family members, HilC and HilD, directly bind and derepress the Salmonella typhimurium hilA promoter. Molecular Microbiology, 40:1289–1299.

132. Ellermeier, C.D., and Slauch, J.M. (2003) RtsA and RtsB coordinately regulate expression of

the invasion and flagellar genes in *Salmonella enterica* serovar Typhimurium. Journal of Bacteriology, 185:5096–5108.

133. Darwin KH, Miller VL. 2000. The putative invasion protein chaperone SicA acts together with InvF to activate the expression of *Salmonella typhimurium* virulence genes. Molecular Microbiology, 35(4):949-959.

134. Lambert MA, Smith SG. 2008. The PagN protein of *Salmonella enterica* serovar Typhimurium is an adhesin and invasin. BMC Microbiology, 8:142.

135. Lambert MA, Smith SG. 2009. The PagN protein mediates invasion via interaction with proteoglycan. FEMS Microbiology Letters 297:209–216.

136. Barthel M, Hapfelmeier S, Quintanilla-Martinez L, Kremer M, Rohde M, et al. 2003. Pretreatment of mice with streptomycin provides a *Salmonella enterica* serovar Typhimurium colitis model that allows analysis of both pathogen and host. Infection and Immunity, 71**:**2839–2858.

137. Hapfelmeier S, Ehrbar K, Stecher B, Barthel M, Kremer M, et al. 2004. Role of the Salmonella pathogenicity island 1 effector proteins SipA, SopB, SopE, and SopE2 in *Salmonella enterica* subspecies 1 serovar Typhimurium colitis in streptomycin-pretreated mice. Infection and Immunity, 72:795–809.

138. Petrone BL, Stringer AM, Wade JT. 2014. Identification of HilD-Regulated Genes in *Salmonella enterica* Serovar Typhimurium. Journal of Bacteriology, 196(5):1094 –1101.

139. Galan JE, Curtiss R. 1989. Cloning and molecular characterization of genes whose products allow *Salmonella* Typhimurium to penetrate tissue culture cells. PNAS*,* 86: 6383–6387.

140. Jiang L, Feng L, Yang B, Zhang W, Wang P, et al. 2017. Signal transduction pathway mediated by the novel regulator LoiA for low oxygen tension induced Salmonella Typhimurium invasion. PLoS Pathogens, 13(6):e1006429.

141. Kingsley RA, Kay S, Connor T, Barquist L, Sait L, et al. 2013. Genome and transcriptome adaptation accompanying emergence of the definitive type 2 host-restricted *Salmonella enterica* serovar Typhimurium pathovar. mBio, 4(5):e00565-13.

142. Miao EA, Leaf IA, Treuting PM, Mao DP, Dors M, et al. 2010. Caspase-1-induced pyroptosis is an innate immune effector mechanism against intracellular bacteria. Nature Immunology, 11:1136–1142.

143. Hersh D, Monack DM, Smith MR, Ghori N, Falkow S, et al. 1999. The Salmonella invasin SipB induces macrophage apoptosis by binding to caspase-1. PNAS, 96:2396–2401.

Acknowledgements

I would like to thank for his endless support to my PhD project my tutor Gabriele Casadei, and Stefano Pongolini, Luca Bolzoni, Erika Scaltriti, Marina Morganti, Melissa Berni, Chiara Bracchi and the other members of the Risk Analysis and Genomic Epidemiology Unit in IZSLER-Parma for providing me their help and assistance throughout the PhD period. I have been lucky to work with them over these years.

I would like to express my gratitude to Robert Kingsley, for giving me the opportunity to work for 5 months on my PhD project in his lab, and to Gaëtan Thilliez, Jennifer Tanner and the other members of Kingsley group for their help during my lab work.

My sincere thanks also to the coordinator of the PhD program in Biotechnology and Bioscience Prof. Simone Ottonello for his assistance during the PhD period.

I would like to thank my PhD colleague Valentina Buffagni, with whom I have shared both stressful and exiting moments throughout these years.

Last I would like to express my deep gratitude to my family and friends for their love and their kind support to all my life.