UNIVERSITÀ DEGLI STUDI DI SASSARI

DIPARTIMENTO DI SCIENZE BIOMEDICHE

DOTTORATO DI RICERCA IN

BIOCHIMICA, BIOLOGIA E BIOTECNOLOGIE MOLECOLARI

XXI CICLO

Coordinatore Prof. Bruno Masala

Structural and Functional Genomic Analysis of the *Salmonella enterica* Host-Restricted Serotype Abortusovis

Tutor:

Prof. Sergio Uzzau

Tesi del

Dr. Massimo Deligios

Anno Accademico 2007-2008

UNIVERSITÀ DEGLI STUDI DI SASSARI

DIPARTIMENTO DI SCIENZE BIOMEDICHE

DOTTORATO DI RICERCA IN

BIOCHIMICA, BIOLOGIA E BIOTECNOLOGIE MOLECOLARI

XXI CICLO

Coordinatore Prof. Bruno Masala

Structural and Functional Genomic Analysis of the *Salmonella enterica* Host-Restricted Serotype Abortusovis

Tutor:

Prof. Sergio Uzzau

Tesi del

Dr. Massimo Deligios

Il Coordinatore:

Prof. Bruno Masala

Anno Accademico 2007-2008

Table of contents

List of tables

List of figures

List of abbreviations

Bp	Base pairs
CDS	Coding sequence
CFU	Colony forming unit
CN	Copy number
DC	Dendritic cell
dNTPs	Deoxynucleoside triphosphates
dsDNA	Double- stranded DNA
EC	Enteric cell
ELISA	Enzyme-linked immunosorbent assay
ETEC	Enterotoxigenic *Escherichia coli*
GC	guanylyl cyclase
MHC	Major histocompatibility complex
OMP	Outer membrane protein
ORF	Open read frame
PCR	Polymerase chain reaction
PEEC	Pathogen-elicited epithelial chemoattractant
PFGE	Pulsed field gel electrophoresis
PMA	Phorbol myristate acetate
RNS	Reactive nitrogen species
ROS	Reactive oxygen species
SAO	*S. enterica* subsp. *enterica* serotype Abortusovis
SAR	*S. enterica* subsp. *arizonae*
SB	*S. bongori*
SCS	*S. enterica* subsp. *enterica* serotype Choleraesuis
SCV	*Salmonella* containing vacuole
SD	*S. enterica* subsp. *enterica* serotype Dublin
SG	*S. enterica* subsp. *enterica* serotype Gallinarum
SHA	*S. enterica* subsp. *enterica* serotype Hadar
SI	*S. enterica* subsp. *enterica* serotype Infantis
Sif	*Salmonella*-induced filaments
SPA	*S. enterica* subsp. *enterica* serotype Paratyphi A
SPI	*Salmonella* pathogenicity island
spp.	Species
Spv	*Salmonella* plasmid virulence
ssp.	Subspecies
STa	*E. coli* heat-stable enterotoxins
STM	*S. enterica* subsp. *enterica* serotype Typhimurium
STY	*S. enterica* subsp. *enterica* serotype Typhi
T3SS	Type III secretion system

ABSTRACT

Abstract

Salmonellae cause diseases both in cold-blooded and in warm-blooded. To date, thousands of serotypes have been identified belonging to *Salmonella enterica* and each of these shows different host range and pathogenicity. Therefore, it is of primary importance to understand the molecular (genetic) basis of host-specificity within serotypes of this species.

In this work, we have performed a structural and functional analysis of the genome of one of such *Salmonella enterica* host-restricted serotypes: serotype Abortusovis (strain SS44). The structural analysis was accomplished throughout the whole genome sequencing using the 454 sequencing technology. We obtained the sequence, analysed it, and compared it with other sequenced genomes. Also, using physical maps and macro-restrictions, we described genome plasticity and structural variation among strains. Plasticity was shown to be high in strains harbouring the IS*1414* insertion sequence. Furthermore, two groups of strains displayed large rearrangements within *Salmonella* Pathogenicity Islands, that are loci otherwise conserved.

Finally, functional aspects have been also valuated in this work. The role of plasmid-encoded effectors has been observed using an experimental model of infection in lambs, and, in vitro, in phagocytic cells (THP-1). This allowed to identify and

characterise a fragment of the SS44 strain virulence plasmid (pSS44) which seems to encode genes with a role in dampering the inflammatory host response.

Chapter 1

INTRODUCTION

Salmonella-host interaction

Serotype host-specificity

Salmonellae are rod-shaped, Gram-negative, facultative intracellular enterobacteria. Salmonelloses are the most common food-borne bacterial diseases in humans and represent a major public health and economical burden worldwide (Grassl and Finlay, 2008).

The genus *Salmonella* consists of two species: *Salmonella bongori* and *Salmonella enterica*. The latter is further subdivided in six subspecies: *S. enterica* subsp. *enterica*, *S. enterica* subsp. *salamae*, *S. enterica* subsp. *arizonae*, *S. enterica* subsp. *diarizonae*, *S. enterica* subsp. *indica*, and *S. enterica* subsp. *houtenae*, or I, II, IIIa, IIIb, IV, and VI, respectively. The subspecies I is the most interesting because it is associated with diseases in warm-blooded animals. So far, over 1500 different serotypes (or serotypes) were identified within the subspecies I, using the Kauffmann-White scheme. All together, these serotypes cause diseases in a wide range of animals (Table 1) (Grimont and Weill, 2007). The symptoms range from mild enteritis to typhoid fever, depending on the serotype and the host; moreover, each serotype is characterised by a specific host range. Certain Salmonellae serotypes, named ubiquitous, infect all warm-blooded animals (such as *S. enterica* serotypes Typhimurium and Enteritidis), causing primarily enteritis. On the other hand, host-adapted serotypes can infect few hosts, mostly causing a clinically relevant disease only in one animal species (e.g. *S. enterica* serotype

Choleraesuis in swine and Dublin in cattle). Moreover, there are serotypes that are host-restricted (e.g. *S. enterica* serotypes Typhi and Abortusovis infecting human and ovine, respectively) causing, as a rule, systemic diseases (Uzzau *et al.*, 2000a). Finally, it is noteworthy that strains of *S. enterica* ser. Typhimurium have been described as host-restricted (Rabsch *et al.*, 2002).

Family	Genus	Species	Subspecies	Serotypes
Enterobacteriaceae	*Salmonella*	*S. enterica*	*enterica* (I)	Typhimurium, Typhi
			salamae (II)	9,46:z:z39
			arizonae (IIIa)	43:z29:-
			diarizonae (IIIb)	6,7:l,v:1,5,7
			houtenae (IV)	21:m,t:-
			indica (VI)	59:z36:-
		S. bongori	(V)	13,22:z39:-

Table 1. The genus *Salmonella*: two species are showed, the six *S. enterica* subspecies and some examples of serotypes for each subspecies.

Host cell access

Salmonellae are usually transmitted through the oral-faecal route. Bacteria are able to overcome gastric acidity, bile salts and other innate immune defences so as to reach the intestinal mucosa in the distal ileum, where preferably enter Peyer's patches M cells (Haraga *et al.*, 2008). There are two other ways to colonise the intestine: via intestinal epithelial cells (enterocytes) and via CD18-expressing phagocytes. Recent findings reveal that these phagocytic cells are probably dendritic cells (DCs), which are ca-

pable to span enterocytes through their dendrites (Grassl and Finlay, 2008; Rescigno and Chieppa, 2005; Rescigno *et al.*, 2001) (Figure 1).

Since enterocytes are not professional phagocytic cells, *Salmonellae* have evolved a sophisticated repertoire of virulence factors that allow them to enter the intestinal epithelium. The most important virulence factors are encoded by the *Salmonella* pathogenicity island 1 (SPI-1) and translocated by the *Salmonella* type III secretion system (T3SS, encoded also by SPI-1). These effectors induce bacterial-mediated endocytosis in non-phagocytic cells by active induction of membrane ruffles that surround bacteria. As *Salmonellae* reach the subephitelial compartment (lamina propria), they can interact with DCs (Biedzka-Sarek and El Skurnik, 2006; Rescigno, 2002; Wick, 2007).

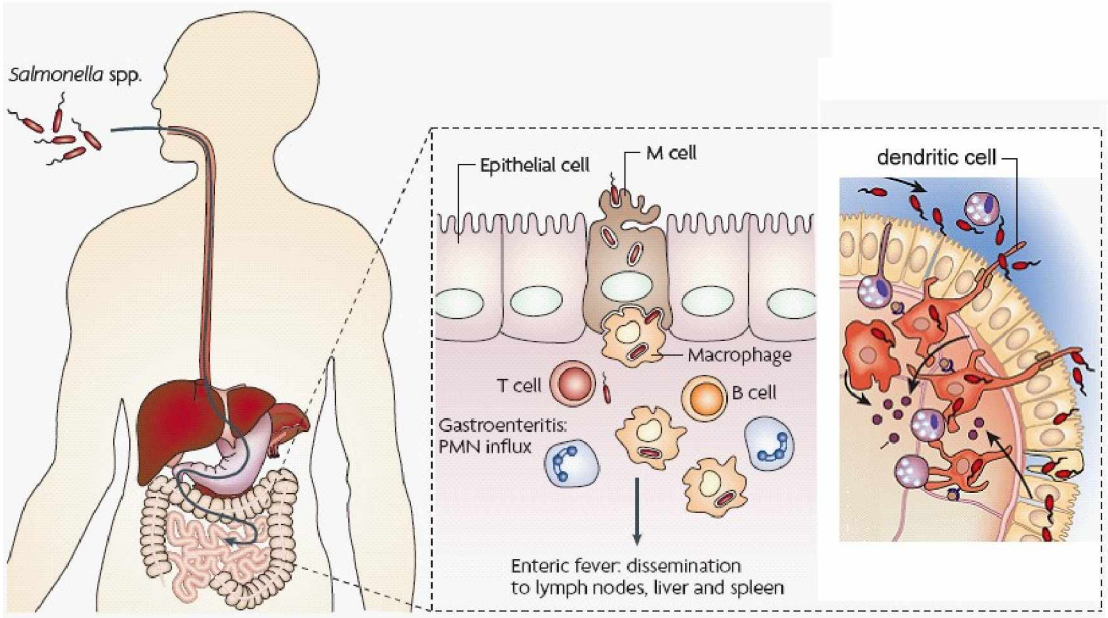

Figure 1. *Salmonellae* colonisation of the host. The figure shows how bacteria translocate through intestinal mucosa: via epithelial, M and dendritic cells (image modified from (Haraga, *et al.*, 2008; Rescigno and Chieppa, 2005).

The type III secretion system

The *Salmonella* genome contains several clusters of genes, called *Salmonella* pathogenicity islands (SPIs), encoding virulence factors. In particularly, SPI-1 comprises genes required for the bacterial entry, while SPI-2 is involved in intracellular survival (in macrophages, DCs and possibly enterocytes). A functional interplay between these two islands has been clearly demonstrated. They both encode type III secretion systems (T3SS-1 and T3SS-2), a molecular syringe consisting of more than 20 proteins spanning the two bacterial membranes (Galan and Wolf-Watz, 2006). This refined apparatus permits the translocation of bacterial factors (called effectors and encoded within the SPI and in other regions) into the host cell cytosol (Figure 2) (Abe *et al.*, 2005; Cornelis, 2006).

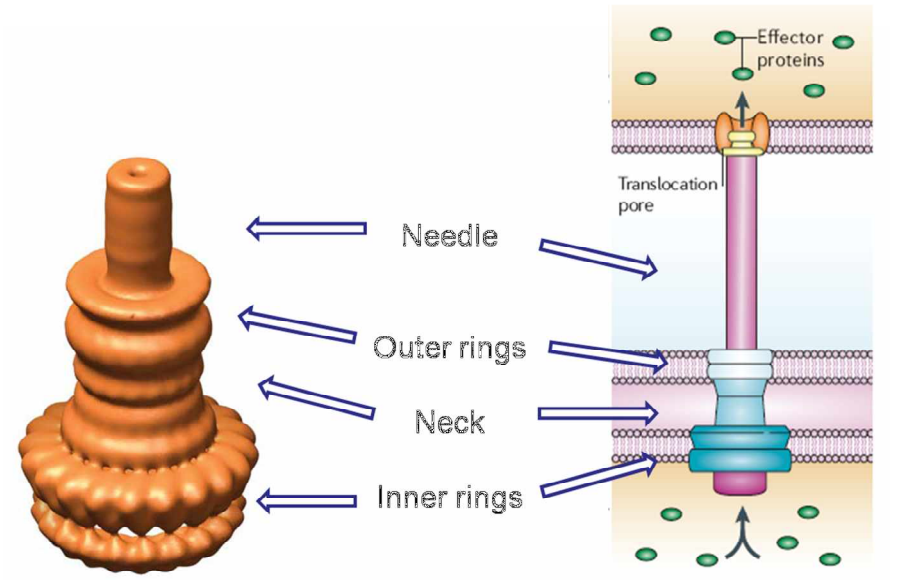

Figure 2. The type III secretion system. In the left the three-dimensional structure of the machinery is showed. The *Salmonella* double membrane and the host membrane are showed in the right. Effectors are directly translocated into host cytosol (image modified from (Cornelis, 2006; Galan and Wolf-Watz, 2006).

The inflammatory response during *Salmonella* infection

Model infection in calves with *S. enterica* serotype Typhimurium results in enteritis. Neutrophils are apparently the main cells involved in *Salmonella*-induced diarrhoea. *Salmonella* serotypes that induce enteritis are able to induce neutrophils influx via transepithelial signalling cell induction (Tukel *et al.*, 2006).

Salmonella infection provokes IL-8 production via NF-κB activation in epithelial cells that, in turns, induces a strong neutrophils influx. This process requires SPI-1 effectors. Furthermore, another factor named pathogen-elicited epithelial chemoattractant (PEEC) has been described to act in concert with IL-8, in order to promote neutrophils migration, in a NF-κB independent manner.

Furthermore, *Salmonella*-induced apoptosis induces inflammation. It is referred that apoptotic macrophage secretes the potent pro-inflammatory IL-1 after *Salmonella* infection (Fink and Cookson, 2007). Therefore, it was coined a new term for this proinflammatory programmed cell death: the pyroptosis.

It is a programmed cell death as apoptosis but, conversely, induces proinflammatory factors release as necrosis (Fink and Cookson, 2005; 2007). *Salmonellae* are able to induce this process via activation of IL-1 β-converting enzyme (caspase-1) that cleaves the inactive precursors of IL-1β and IL-18 into mature cytokines.

SPI-1 T3SS and flagellin are needful for trigger pyroptosis in macrophages and this process is shared by different serotypes: broad range and host specific salmonellae are equally cytotoxic for murine macrophages. Also, caspase-1 activation is involved in *Salmonella* infected DCs pyroptosis.

Pyroptosis mechanism is quite different from apoptosis; in fact, the most important effector is caspase-1 rather than caspase-3. Furthermore, caspase-6 and caspase-8 and mitochondria released cytochrome-C are not implicated. Conversely, DNA degradation occurs during both events, even if pathways are different.

Systemic spread

Once *Salmonella* cross epithelium, it encounters DCs. These bacteria have developed efficient mechanisms to escape host defences. First, *Salmonella* can induce pyroptosis via SPI-1, so they can evade from phagocytic host cells, such as DCs. Also, through SPI-2, bacteria are able to survive within host cells and spread in the organism.

DCs are bone-marrow-derived phagocytic cells (Biedzka-Sarek and El Skurnik, 2006). They present long cellular processes called dendrites. When the epithelium undergoes an infection, enterocytes secrete CCL-20 and CX_3CL-1 chemokines that induce a DCs influx in the subepithelial region. CX_3CL-1 stimulates DCs for dendrites extension. They cross the epithelium and sample antigens such as *Salmonella*. At this point, *Salmonella* infection might have different outcomes. In some cases, DCs and macrophages may undergo *Salmonella*-induced pyroptosis or necrosis. Otherwise, DCs can

acquire antigenic material by remnants and stimulate T cells. Also, DCs pyroptosis provokes neutrophils activation and new DCs influx. Other DCs, which do not undergo the apoptosis, can migrate to secondary lymphoid organs. Meanwhile, DCs produce a number of surface molecules, such as MHC-I, MHC-II, CD40, CD80, CD86 and CD54. Therefore, DCs migrate towards secondary lymphoid organs where they present *Salmonella*-derived peptides to naïve $CD4^+$ or to $CD8^+$ T cells. Activated T cells then migrate to the site of inflammation, while activated DCs produce pro-inflammatory cytokines TNF-α and IL-12. Finally, intracellularly viable *Salmonella* may also elicit nitric oxide (NO) by DCs but slow phagosome maturation protects bacteria. Also, NO stops T cells proliferation.

Once *Salmonella* cross the intestinal mucosa, they reach lymphoid organs, such as spleen and liver. Bacteria replicate within macrophages and then shed in the bloodstream causing septic shock and multiple organ failure.

DCs are powerful APCs, antigens internalised are presented to T cells via major histocompatibility complex class I (MHC-I) or II (MHC-II). Endogenous antigens are cleaved by proteasomes in the cytoplasm and then loaded in the MHC-I. Therefore, MHC-I-peptide complexes are exposed in the plasma membrane and can be presented to $CD8^+$ T cells. Conversely, exogenous antigens (such as bacteria) are processed in phagolysosomes. Peptides are loaded into MHC-II and, after their transport to the cell surface, presented to $CD4^+$ T cells (T-helper cells). However, occasionally peptides exogenous can be associated to the MHC-I and peptides endogenous can be loaded into MHC-II.

Dendritic cells maturation

TLRs are receptors that interact with bacterial ligands, such as LPS and flagellin (Artis, 2008; Guo *et al.*, 2007; Iliev *et al.*, 2007). Following their activation, DCs migrate to lymphoid organs and simultaneously they mature. The first sign of the maturation is the loss of phagocytic activity. It is much interesting how *Salmonella* can affect the DC maturation. *Salmonella* are able to upregulate MHCs and co-stimulatory molecules (CD86, CD54 and CD40) expression.

When *Salmonella* infect the host, DCs secrete various molecules, such as IL-12. IL-12 stimulates natural killer cells, macrophages, T cells and DCs themselves for IFN-γ production. IFN-γ is crucial for the bactericidal activity induction in macrophages and T-helper (Th1) cells development. IL-18 is as important as IL-12 for IFN-γ induction. SipB SPI-1 effector seems to be involved in IL-18 secretion.

The *Salmonella* Genome

Salmonellae genomes are constituted by a circular chromosome (about 5 million bases pairs) and usually one to three plasmids. To date, over ten *Salmonella* strains have been completely sequenced and a number are in finishing phase (Chiu *et al.*, 2005; Deng *et al.*, 2003; McClelland *et al.*, 2004; McClelland *et al.*, 2001; Parkhill *et al.*, 2001). These data have been used to redefine the relationship between different *Salmonellae* and, in the near future, comparative analysis will allow new diagnostic and therapeutic approaches.

Genome plasticity

Genetic similarity among *Salmonellae* species and serotypes was recognised firstly by reassociation methods, then by physical mapping and finally by whole genome sequencing. In the "Pre-Genomic Era" reassociation of DNA was used to compare DNA sequences from different organisms, since, the kinetics of the reassociation process depends on DNA similary (Crosa *et al.*, 1973). More recently, pulse-field gel electrophoresis (PFGE) following I-CeuI partial digestion was used to obtain physical

maps (S. L. Liu *et al.*, 1993; W. Q. Liu *et al.*, 2007). This enzyme cuts all seven *rrl* genes (harboured in the *rrn* operons) present in enteric bacteria, showing the order of these fragments. Also, PCR amplification designed on fragments junction, allows to define the orientation of these seven regions.

Even though some serotypes show rearrangements, their genomes can be considered conserved. Indeed, during the bacterial replication, duplications and inversions are common but selective pressure seems to remove these rearrangements. It has been showed that genetic mechanisms, such as genome balance and gene dosage, prevent large genomic variations. Genome balance preserves length of the replichores between the oriC and Ter sites. Any alteration of the genomic structure ought to be compatible with this balance. Indeed, all strains analysed have balanced arrangement. Besides, the different genes have evolved promoter strength according to their distance from oriC or Ter, since during replication the expression near these sites is higher. Hence, also gene dosage prevents high levels of gene arrangements.

It is known that in some species (i.e. *Yersinia pestis*), insertion sequences can mediate large recombination; however, the most of recombination in serotype Typhi occurred between *rrn* operons.

Analysis of genomic structure of *Salmonellae* demonstrated that ubiquitous serotypes harbours more conserved genomes. Conversely, serotypes Gallinarum, Typhi and Paratyphi C (host-restricted) show rearrangements (Holt *et al.*, 2008; Hughes, 2000; Kothapalli *et al.*, 2005; Gui-Rong Liu *et al.*, 2006; W. Q. Liu*, et al.*, 2007).

Mobile genetic elements

Salmonella genomes contain several loci that are unstable, such as plasmids, bacteriophages and insertion sequences, and other DNA regions. These regions, even horizontally transferred and linked to the pathogenicity, are permanently included in the genomic backbone (pathogenicity islands).

Plasmids

Many *Salmonellae* harbour plasmids, some of which can be transferred by conjugation (Camacho and Casadesus, 2005; Camacho *et al.*, 2005a; Camacho *et al.*, 2005b; Chu *et al.*, 2002). Salmonella plasmids are involved in virulence and in drug resistance (Rychlik *et al.*, 2006). Within subspecies I, probably at least ten serotypes contain F-type virulence plasmids: *S. enterica* serotypes Typhisuis, Typhimurium, Paratyphi C, Enteritidis, Dublin, Choleraesuis, Abortusovis, Abortusequi, Gallinarum-Pullorum and Sendai (Baumler *et al.*, 1998; Chu and Chiu, 2006; Uzzau *et al.*, 2005). Most plasmids are reported to be mobilisable, but only few are able to undergo conjugation, such as plasmid pSLT in *S. enterica* serotype Typhimurium. So far, it was published the complete sequences of only few virulence plasmids (Table 8), even though heteroduplex analysis indicates that the majority are closely related. All virulence plasmids harbour a common DNA region encoding *spv* operon, whereas in other subspecies this region is

located on the chromosome (Boyd and HartI, 1998). As discussed above, SpvB is involved in cell death via pyropoptosis (Chu *et al.*, 2008; Guiney and Lesnick, 2005).

Often, *Salmonellae* drug resistance is encoded by plasmids of H incompatibility group (IncHI1). These plasmids ensure resistance versus different antibiotics and are transmissible between members of *Enterobacteriaceae* and other Gram-negative bacteria. Furthermore, virulence plasmids can be associated also with antibiotic resistance genes, such as OU plasmids in serotype Choleraesuis.

Finally, different serotypes sometimes carry cryptic plasmids which are supposed to be involved in pathogenicity even not containing any virulence genes.

Bacteriophages

Salmonellae harbour several bacteriophage genes within the chromosome. Bacteriophages carry often additional genes (cargo genes) that can be useful for hosts. In some cases phages are composed by few (less than five) genes, generally including transposases and derive by following phage excision event. In others, they are prophages or phages remnants composed by large clusters of genes. The two lamboid-like prophages Gifsy 1 and Gifsy 2 affect the virulence of their hosts. Therefore, the curing of Gifsy 2 reduces the virulence of serotype Typhimurium for mice by up to 100 fold (Figueroa-Bossi and Bossi, 1999). Gifsy phages include cargo genes such us *sodCI*, encoding a superoxide dismutase involved in resistance to the macrophage oxidative burst. Also, Gifsy 1 encodes *gipA* (associated with intestinal colonisation) and *sseI*

(T3SS SPI-2 secreted). Another example is *sopE* phage, homologue with bacteriophage P2 and encoding *sopE*.

Salmonella pathogenicity islands

As discussed above, *Salmonella* pathogenicity islands (SPIs) are large regions of DNA associated with virulence traits and probably have been horizontally transferred into the genome. This transfer allows the rapid acquisition of complex virulence functions from other pathogens and the extension of the host range. Pathogenicity islands are often genetically unstable and harbour genetic elements involved in DNA-mobility, such us integrases and transposases. Moreover, base composition of their DNA (e.g. % G + C) is generally different from that of the core genome (Wilson and Nickerson, 2006).

To date, several SPIs have been identified, but only few are present in all serotypes analysed. In order to check the SPIs presence, a number of microarrays analyses have been performed. Nowadays, with availability of several genome sequences, new potentially horizontally transferred regions can be identified. For example, ten potential SPIs are present in *S. enterica* serotype Typhi CT18, but only SPI-1 to 5 have been well characterised in Typhimurium. Even though SPIs are almost stabile in the genome, variations have been reported between the serotypes (Amavisit *et al.*, 2003).

The *Salmonella* pathogenicity island I

As described above, SPI-1 encodes structural proteins of T3SS and effectors which are translocated across this machinery. Nevertheless, not all effectors are encoded by SPI-1; in fact *sopA, sopB, sopD, sopE* and *sopE2* are encoded by independent loci. The island is about 40 kb in size and is present in all *Salmonella* genus and is almost conserved, also there are not genetic elements associated with DNA instability.

The activation of the SPI-1 genes is strictly regulated by the protein HilA, encoded itself by SPI-1 (Altier, 2005; Boddicker *et al.*, 2003; Dieye *et al.*, 2007; Jones, 2005). This protein controls also other SPIs encoded genes, such as those of SPI-4 and SPI-5. Activation of SPI-1 depends on several conditions such as pH and ppGpp (Perez and Groisman, 2007; Thompson *et al.*, 2006). Furthermore, it has been demonstrated that PhoP/PhoQ two-component system represses the expression of HilA (Aguirre *et al.*, 2006; Groisman, 2001; Prost and Miller, 2008).

The expression of the T3SS-1 and the effectors are induced promptly as *Salmonella* cells reach the distal small intestine. Upon host cell contact, bacteria deliver the effectors into the cell, interfering with a broad variety of cellular processes (Ellermeier and Slauch, 2007; Schlumberger and Hardt, 2006; Suarez and Russmann, 1998; Winnen *et al.*, 2008).

At least 13 proteins have been identified to be secreted by the T3SS-1 in *Salmonella* serotypes: AvrA, SipA, SipB, SipC, SipD, SIrP, SopA, SopB/SigD, SopD, SopE, SopE2, SptP and SspH 1 (Ly and Casanova, 2007). Six of them have shown to control actin cytoskeleton dynamics.

It is reported that AvrA is a deubiquitinase that stabilises the tight junctions in epithelial cells during the inflammation (Liao *et al.*, 2008; Ye *et al.*, 2007). SipC composes part of the bacterial translocon. It contains two domains spanning the host plasma membrane. The C-terminal cytoplasmic domain nucleates the assembly of actin filaments; it is also needed for effector protein translocation. As reported elsewhere, these two functions are dissociable. SipA enhances actin filaments assembly and inhibits their disassembly; moreover, it disrupts the tight junctions (Lilic *et al.*, 2003). SopE and SopE2 are guanine nucleotide exchange factors (GEFs) that activate Rac and Cdc42. These host proteins, when active, trigger the membrane ruffling. Another effector protein is SopB, which triggers actin rearrangements. It seems that SopB activates RhoG.

Therefore, SipA and SipC can produce small F-actin foci at site of bacteria-host cell contact, but they cannot induce ruffling themselves since they need SopE, SopE2 and SopB. A Δ*sopEsopE2sopB* triple mutant defects in ruffling and cannot induce invasion in the intestinal mucosa.

The bacterial internalisation is completed with the phagosome closure, named *Salmonella* containing vacuole (SCV). This process requires, at least, the SopB activity. After bacterial entry, the host cell regains its cytoskeleton structure and apical microvilli. This re-establishment is mediated by secreted effector SptP (Lin *et al.*, 2003). This protein contains two different catalytic portions: a RhoGAP domain and a tyrosine phosphatase. The first deactivates Rac and Cdc42 resulting in a reduction of actin assembly. The latter is also involved in cytoskeletal recovery after bacterial entry. Thus

SopE and SptP are antagonists. While they are delivered simultaneously, their half-life is strikingly different: SptP persists several hours in the host cell, while SopE is degraded 30 minutes after the infection (Kubori and Galan, 2003).

The *Salmonella* pathogenicity island II

Conversely to SPI-1, SPI-2 is less conserved in *Salmonellae*. It can be considered an evolutionary acquisition in *Salmonellae* that infect warm-blooded animals. Also SPI-2 is 40 kb in size and encodes a T3SS which delivers effectors across the SCV. Moreover, SPI-2 encodes tetrathionate reductase (Ttr) and only this part is the present in *S. bongori*.

When bacteria enter host cell, they are enclosed into the membrane-bound compartment SCV (Drecktrah *et al.*, 2007; Gorvel and Meresse, 2001). SCV presents only some late-endosomal markers but not the most important lysosomal receptors; this fact suggests that it does not fuse directly with lysosomes, but follows a proper fate that avoids bacterial killing (Steele-Mortimer, 2008). Few hours after infection, SCVs show long tubular structures called *Salmonella*-induced filaments (SIF). To date their role is unclear, but their involvement in *Salmonella* pathogenesis is unquestionable since a number of virulence determinants promote their organisation (Abrahams and Hensel, 2006).

While SPI-1 is involved in host cell entering, the intracellular survival is mediated by genes encoded in the SPI-2, recent studies propose that SPI-2-encoded genes may be expressed in the intestinal epithelium before access (Brown *et al.*, 2005).

SPI-2 encodes both T3SS-2 and regulatory system SsrAB that controls the expression of the SPI-2 regulon. SrrAB is controlled itself by a global regulatory system. It is interesting to note that SPI-2 encodes only few effectors delivered by T3SS-2, while the majority are encoded in additional loci scattered throughout the genome. Such physical organisation indicates the ongoing acquisition of new effectors correlated to the SPI-2 (Lober *et al.*, 2006).

A number of T3SS-2 effectors are associated with endosomal membrane. These include SseF, SseG, SseJ, SifA, SifB, PipB, PipB2 and SopD2. In fact, sequence analysis has shown hydrophobic domains that could mediate membrane association (Waterman and Holden, 2003).

Other effector proteins interfere with cellular transport processes, such as SpiC. SpiC binds TassC which is probably involved in vesicular trafficking and/or fusion. Moreover, SpiC is required for *in vitro* secretion of other SPI-2 effectors (such as SseB, SseC and SseD). Thereafter the *spiC* mutant fails not only to produce SpiC protein *per se* but also probably to deliver all SPI-2 effectors.

The effectors SseF and SseG bind microtubules. They are likely transmembrane proteins and they might have a role in vacuolar membrane dynamics.

So far, the best characterised T3SS-2 effector is SifA. This protein controls the position of the SCV preventing the excessive recruitment of the motor-protein kinesin to the SCV. In fact, it targets SKIP (SifA and kinesin-interacting protein) which down-regulates the kinesin activity and hence the association between kinesin and SCV. Also,

SifA and SseG seem to control the motor-protein dynein recruitment to the SCV. Furthermore, kinesin and dinein act for SIF formation.

SPI-2 effectors are also involved in the avoidance of reactive oxygen and nitrogen species (ROS and RNS). High levels of ROS are prevented because altered actin structure in the SCV avoids the function of a NADPH-oxidase (phox) in the SCV membrane. Regarding RNS, it was described that T3SS-2 inhibits iNOS functional, even though specific effectors are not already identified.

Whereas T3SS-1 and T3SS-2 secreted genes, and virulence plasmid encoded genes, are needful for systemic spreading; only SPI-1 is required for inducing enteritis. However, it was identified another effector, SopB, encoded by SPI-5, linked to the pathogenesis of diarrhoea. SopB is an inositol phosphate phosphatase that increases chloride secretion. Nevertheless, this is not sufficient for inducing the diarrhoea because it is associated with a strong neutrophils influx. Furthermore, SopB is involved in cytokine expression, such as Akt activation.

SopD, regulated by SPI-1, is another effector that strengthens SopB in enteritis induction, whereas SopA is the one that elicits transepithelial migration of neutrophils. It was demonstrated that SipA, SopA, SopB, SopD and SopE2 act in concert to induce diarrhoea (Ehrbar *et al.*, 2003; Grassl and Finlay, 2008; Santos *et al.*, 2003; Zhang *et al.*, 2003; Zhang *et al.*, 2002).

Other *Salmonella* pathogenicity islands

SPI-3 is conserved between serotypes Typhi and Typhimurium but analysis in other *Salmonellae* revealed deletions and insertions. The main virulence factor encoded by SPI-3 is high affinity magnesium transport system MgtCB. This factor is important for intracellular survival of Salmonella and the encoding region is much conserved between the genus *Salmonella* (Blanc-Potard *et al.*, 1999).

SPI-4 is 27 kb in size and is constituted by *siiABCDEF* operon. SiiC, SiiD and SiiF form a type I secretion system for secretion of SiiE, a huge (about 600 kDa) protein associated with the bovine intestinal colonisation (Kiss *et al.*, 2007).

SPI-5 is 7.6 kb in size and harbour *sopB* gene. This gene is present in all Salmonellae whereas *pipAB* part is less conserved. SopB is delivered by T3SS-1 instead PipB by T3SS-2 (Wood *et al.*, 1998).

SPI-6 locus is less conserved and functions of few genes are known. SPI-6 encodes *saf fimbrial* operon that could potentially be involved to virulence in *Salmonella* (Folkesson *et al.,* 1999).

SPI-7 is the largest SPI (133 kb) and this is identified only in serotypes Typhi, Dublin and Paratyphi C. The region is composed by individually acquired elements and *pil*, *tra* and *sam* genes derive by conjugative plasmid or transposon. The *sopE* phage is probably recently acquired because is not present in serotypes Dublin and Paratyphi C, besides it is harboured by isolates that lack SPI-7. The virulence factors encoded by SPI-7 are the *viaB* gene locus (encoding Vi exopolysaccaride capsule), the sopE prophage (effector protein translocated by T3SS-1) and a gene cluster encoding type IVB pili (in-

volved in invasion of epithelial cells by serotype Typhi) (Lee *et al.*, 2006; Pickard *et al.*, 2003).

There are other three pathogenicity islands identified in serotype Typhi: SPI-8, encoding bacteriocins; SPI-9, encoding a T1SS and a large RTX-like protein; SPI-10, encoding the Sef fimbriae.

Salmonella enterica serotype Abortusovis

S. enterica serotype Abortusovis (antigenic formula O:4,12:c:1,6) is a host-restricted *Salmonella* since it has been isolated only from ovine source. Serotype Abortusovis is the main responsible of ovine abortion in Europe and western Asia. Infection causes abortion and mortality in newborn lambs, whereas natural exposition in adult sheep does not produce symptoms. After abortion, *Salmonellae* can be isolated from placental and foetal tissues, which are the principal sites of multiplication. Vaginal discharges following abortion can contaminate food and spread the infection to other sheep (Bacciu *et al.*, 2004).

Studies conduced in the murine study model of infection in Balb/C show that the serotype Abortusovis is able to colonise the ileum but few bacteria reach Peyer's patches with respect to the serotype Typhimurium infection. Ovine adaptation has favoured systemic asymptomatic infection in adult sheep probably losing gene functions involved in enteritis induction. Again, host can be implicated in process of adaptation. In fact, lambs experience enteritis but not the adult sheep. The explanation could be that the lymphopoiesis in foetal Peyer's patches in the jejunum is very intense whereas the process involves in ewes, which are only few follicles detectable after 18 months of age. Hence, Abortusovis probably infects ewe's intestine transiently while during preg-

nancy bacteria are attracted in foetal follicles, producing intensive colonisation. This induces abortion or delivery of an ill lamb.

Abortusovis harbours a virulence plasmid ranging in size from 50 to 67 kb in all strains analysed. Experiments performed with a strain of Abortusovis SS44 cured (without virulence plasmid, named SU40) suggest that the virulence plasmid affects systemic infection after oral inoculation but not after intraperitoneal administration in the mouse model (Uzzau *et al.*, 2000b). Also, SU40 strain seems to induce a serious inflammation reaction when inoculated subcutaneously whereas SS44 strain wild type does not produced high levels of swelling (Uzzau, *et al.*, data not published).

Furthermore, serotype Abortusovis carries Gifsy-2-related cryptic prophage, named Gifsy-2AO. Deletion of Gifsy-2AO does not affect intestinal invasion but is associated with attenuation of virulence in the systemic phase of lamb infection. The prophage lacks a 6 kb region of serotype Typhimurium Gifsy-2, that is replaced by an insertion sequence (IS) element named IS*1414* (Bacciu, *et al.*, 2004).

IS*1414* is described initially in enterotoxigenic *Escherichia coli* (ETEC) strain 27D and is constituted by two overlapping genes: *tnpA*, encoding a transposase and in a +1 reading frame, *astA*, encoding a heat-stable enterotoxin EAST1 (McVeigh *et al.*, 2000). EAST1 is very similar to *E. coli* STa (heat-StableToxin a) enterotoxin, that binds guanylyl cyclase C (GC-C), inducing an increase of the concentration of intracellular cGMP (cyclic Guanosine Mono Phosphate) (Park *et al.*, 2000; Veilleux *et al.*, 2008). However, EAST1 is immunologically different by STa since no cross-neutralisation of EAST1 was observed using polyclonal anti-STa antibodies (Savarino *et al.*, 1991).

Strains of Abortusovis harbour generally several copies of IS*1414*, even though strains without this insertion sequence are described. To date, there are not evidence of other *Salmonella* serotypes carrying IS*1414*, with the exception of some serotype Agona strains (Paiva de Sousa and Dubreuil, 2001).

To valuate the heterogeneity of different strains within this serotype, pulsed field gel electrophoresis (PFGE) and IS*200* typing has been performed (Dionisi *et al.*, 2006; Nikbakht *et al.*, 2002; Schiaffino *et al.*, 1996; Valdezate *et al.*, 2007).

Genomics and sequencing technology

The genomics era began in 1995 with the publication of the complete genome of *Haemophilus influenzae* (Fleischmann *et al.*, 1995). The boost over the following decade has had a huge impact on microbial research. Every year, new genomes are added and today hundred of sequences are available. By now, it is possible to obtain a complete microbial genome in few months at ever lower cost. This deal of data allows matching the sequences of related microorganisms (comparative genomics). *In silico* whole-genome analysis can provide to understand pathogenesis of bacteria i.e. identifying the presence of pathogenicity islands and other regions horizontally acquired. Moreover, genome sequence analysis allows deciphering bacterial pathogenesis evolution.

Next-generation sequencing

In 1977, Sanger and Coulson improved a new method for rapid determination of DNA sequence (Sanger *et al.*, 1977). This method has been the state of art for DNA sequencing for 30 years. However, in the last years have appeared other technologies

termed 'next-generation' DNA sequencing, which are substituting the Sanger method (Hall, 2007; Service, 2006).

So far, there are two kinds of sequencing approaches: technologies with amplification or with single-molecule analysis. In the first case, DNA is amplified to obtain an amount of molecules adequate to be analysed. In the Sanger method, this step is performed usually via growth of clones harbouring a plasmid with the sequence. Technologies such us 454 sequencing (Margulies *et al.*, 2005), SOLiD sequencing and Illumina exploit high-throughput DNA amplification (Shendure *et al.*, 2005). Conversely, within two years will be commercialised two new technologies (next-next-generation) based on single molecule analysis (VisiGen and Helicos) (Harris *et al.*, 2008) (Table 2).

Nevertheless, Sanger method is even the most suitable for de novo sequencing because of the longer sequences achieved. However, other systems are developing and probably, between few years substitute completely the Sanger technology.

Technology	
454	www.454.com
SOLiD	www.appliedbiosystems.com
Illumina	www.illumina.com
VisiGen	www.visigenbio.com
Helicos	www.helicosbio.com

Table 2. The new sequencing technologies.

Chapter 2

OBJECTIVES

The main objectives of this research are:

- To perform a structural analysis of a *Salmonella enterica* host-restricted (serotype Abortusovis strain SS44) genome by means of whole genome sequencing, restriction analysis and bioinformatics methods.
- To compare serotype Abortusovis genome sequences with that of other *Salmonella* serotypes already sequenced.
- To investigate the possible role of genetic loci and their correlation with host inflammatory response.

Chapter 3

METHODS

Strains

S. enterica ser. Abortusovis (SAO) strains used in this work belong to a wide collection stored in the laboratories of the Department of Biomedical Sciences (University of Sassari) .The collection includes IS*1414* positive and negative strains originated from various Europe and Asia continents. (Table 3)

Strain #	IS*1414*	Source
SS44	+	Sardinia, Italy
SU40	+	Sardinia, Italy (plasmid-cured SS44)
SSM0041	-	Former USSR, probably Kazakhstan
SSM0071	+	Albania
SSM2026	-	Iran
SSM2027	-	Iran
SSM2028	-	Iran
SSM2029	-	Iran
SSM2030	-	Iran
SSM2031	-	Iran
SSM2032	-	Iran
SSM2035	-	Iran
SSM2037	-	Iran
SSM2044	+	Iran
SSM2045	+	Iran
SSM1379	+	France
SSM1663	+	Sardinia, Italy
SSM1670	+	Sardinia, Italy
SSM3253	+	United Kingdom

Table 3. SAO strains used in this work. For each one is indicated the IS*1414* presence (test by *astA* gene PCR amplification) and the source.

Recombinant strains

Previously, pSS44 (SS44 virulence plasmid) was extracted and digested with *Hin*dIII. Each fragment has been cloned in a pJW4303 vector and each plasmid obtained was used to transform *E. coli* DH5α competent cells. The group of recombinant strains has been stored in our laboratory. Then, we transferred each plasmid from *E. coli* stains to a cured SS44 (SU40) recipient by electroporation (Ausubel, 2001). Briefly, plasmids were been extracted with Qiagen Plasmid Mini Kit and then used to transform SU40 competent cells (Table 4). Each recombinant harbouring a different pSS44 fragment was tested for rate growth and motility. Finally, some recombinants are selected for infection experiments.

Strain # donator	Strain # SU40	Plasmid cloned	Fragment size of pSS44
SSM3189	SSM4329	pDB117	3000
SSM3190	SSM4318	pDB106	2600
SSM3191	SSM4319	pDB107	2600
SSM3192	SSM4320	pDB108	7000
SSM3193	SSM4321	pDB109	3000
SSM3194	SSM4322	pDB110	7000
SSM3195	SSM4323	pDB111	>10000
SSM3196	SSM4324	pDB112	2600
SSM3197	SSM4325	pDB113	2600
SSM3198	SSM4326	pDB114	>10000
SSM3199	SSM4327	pDB115	7000
SSM3200	SSM4328	pDB116	7000+500

Table 4. SU40-derival strains containing a pSS44 fragment cloned into pJW4303. For each DH5α strain (donator) is indicated: the corresponding SU40-derivatives, the plasmid harboured and the estimated size of cloned pSS44 fragment.

Polymerase chain reaction (PCR)

The PCRs were accomplished using *Taq* DNA Polymerase (Invitrogen) 1.25 U, 1x PCR buffer minus Mg, $MgCl_2$ 1.5 mM, dNTP mixture 0.2 mM each, primer mix 0.2 mM each and DNA template prepared using microwave treatment (Ran He, 2005). We have designed primers for different PCRs (Table 5). Amplicons have been separated in 1% agarose gel in TAE buffer with ladder 1 Kb Plus DNA Marker (Invitrogen).

Reaction	Size	Primer #	Sequence
astA	401 bp	02-20	GCGAAGTTCTGGCTCAATGT
		02-21	AGCGACTCGATGGCATTCGT
avrA	385 bp	02-43	AAGTATGCTAAGTCCTACGA
		02-44	GAACAAAATCACCGATGTCT
pSAO0090	769 bp	08-43	GGAAAGGATTCTACCGTACTC
		08-44	AATAGGCACCACTTTCATTG
yagAB	509 bp	08-47	CACTTCTGCGCAGTATGTAA
		08-50	TGCCGAACTGAAAGACGCCT
IS*1414* outward*	-	02-62	TAGCCAGAGTCTGTAACTGT
		02-63	TATTATCGAGTTCGGTGACC

Table 5. Primers used in this study. The amplicon size is reported for each reaction. *These primers are designed outward in IS*1414* sequence and they have been employed in combination with other primers.

Sanger sequencing

Some DNA regions have been sequenced using standard Sanger methods, i.e. gaps between sequences known and plasmidic inserts. In the first case, primers have

been designed in regions flanking the gaps (Table 6) and PCR amplification was performed (see above). We designed also primers in pJW4303 to sequence the insert. Amplicons or plasmids extracted were sent to BMR Genomics (www.bmr-genomics.it) with each primer to be sequenced via Applied Biosystems 3730xl DNA Analyzer.

Reaction	Size	Primer #	Sequence
IT01	146 bp	08-21	AGGATTCTGGTTACCCGATA
		08-22	CAGGACATCCCATTAAAGCT
IT04	1540 bp	08-23	TACTGACTGCCCGGAATTCC
		08-24	GGTCTGCCAATATGACGATG
IT07	1392 bp	08-25	AAGTGGAGTGGTAATATGCG
		08-26	GCAGAGCAACCATGAAGGGT
IN01	-	08-01	CTGACGCCGCTGATAAGCT
		08-32	GTGGTACTGGAAGGCAACAT
IN02	2125 bp	08-03	TGCTCTTCCAGCAGCTGTAT
		08-04	CGTCTGTATACCATTGAGAA
IN04	-	08-31	TAGCACCTTTCAGGCCATAA
		03-34	TGAAATGAGTGCTTATCAG
IN-SPI-1	1600 bp*	08-29	CACCGAAACGATCGGATATA
		08-30	CCGGGATTCACGGAACAATA
pJW4303 insert	-	08-19	GACTGTTCCTTTCCATGGGT
		08-20	CCTCTCTTCATTGCATCCAT

Table 6. Primer designed for gap closure and plasmid sequencing. The size of amplicons is determined after sequencing, if there was not amplification it is not reported. *Data estimated from electrophoresis because there was not amplification in SS44.

Restriction analysis

The virulence plasmids of SAO strains have been extracted via Qiagen Plasmid Midi Kit. A single colony has been picked and bacteria were incubated over night shaking at 37°C in 100 mL of LB broth. The amount of plasmidic DNA has been estimated via electrophoretic separation with High DNA Mass Ladder (Invitrogen). The DNA was digested with *Hin*dIII, *Nco*I, *Bam*HI and *Eco*RI (Invitrogen) in a proper concentration and fragments were separated in a 0.5 % agarose gel in TAE buffer for about 4 hours. Molecular weights are estimated using 1 Kb Plus DNA and TrackIt™ λ DNA/*Hin*dIII Fragments (Invitrogen).

Pulsed-field gel electrophoresis

The SAO strains were typed by pulsed-field gel electrophoresis (PFGE) according to a standardised protocol described elsewhere (Peters *et al.*, 2003). Briefly, after cell lysis by proteinase K, genomic DNA plugs were digested with 24 U of *Xba*I (Promega) and separated on a 1% agarose gel (Agarose LE, Roche) using the CHEF-DR III System (Bio-Rad). Electrophoresis conditions were: run for 22 h at 180 V, with a pulse time of 2-64 s. A *S. enterica* ser. Braenderup strain (kindly supplied by Ida Luzzi from Istituto Superiore di Sanità, Rome) has been chosen as a molecular reference marker.

454 Sequencing

Genomic DNA preparations of the SS44 strain were carried out as described elsewhere (Ausubel, 2001). DNA has been handled according to the manufacturer's instructions to be analysed using 454 Life Sciences GS-20 sequencer and GS-FLX sequencer (Roche, www.roche-applied-science.com). The instruments were used by coworkers of dr. de Bellis (CNR-ITB, Milan). The assembly was accomplished with GS *De Novo* Assembler Software and GS Reference Mapper Software (using published sequences of other *Salmonellae*, see Table 7). As described in Results, paired-end analysis has been performed with the kit provided by the manufacturer.

Bioinformatics

Below are reported the *Salmonella* genomes completely sequenced (Table 7), the respective strain and plasmid sequences (Table 8), used in bioinformatics analysis.

The *in silico* analysis were produced using several computer software. Reads sequences are managed via remote access at the CNR-ITB computers using PuTTY (www.putty.org) and WinSCP (www.winscp.net) software. Easier elaborations, such as contigs and scaffolds analysis, have been performed locally with Vector NTI 10.3.0 (Invitrogen) and Artemis Comparison Tool (The Sanger Institute). The ABI chromatogram files received by BMR Genomics were opened using FinchTV Version 1.4.0 (Geospiza).

Also, web tools were employed to other analysis, such as NCBI portal (www.ncbi.nlm.nih.gov), the Wellcome Trust Sanger Institute web site, where we

found several *Salmonellae* genome sequences (www.sanger.ac.uk) and CGView Server (http://stothard.afns.ualberta.ca/cgview_server/index.html) to display the provisional pSS44 sequence.

Species	Serotype	Strain	Size (bp)	Source
S. bongori		12419	4460000	Sanger*
S. enterica ssp. I	Typhimurium LT2	SGSC1412	4857432	NCBI: NC_003197
	Typhi	CT18	4809037	NCBI: NC_003198
	Choleraesuis	SC-B67	4755700	NCBI: NC_006905
	Paratyphi A	ATCC9150	4585229	NCBI: NC_006511
	Enteritidis	P125109	4686000	Sanger
	Gallinarum	287/91	4747000	Sanger
	Hadar		4797000	Sanger
	Infantis		4711000	Sanger
S. enterica ssp. IIIa	62:z4,z23:--	CDC346-86	4600800	NCBI: NC_010067

Table 7. Sequenced genomes used in comparison. *FastA sequences have been retrieved in www.sanger.ac.uk/Projects/Salmonella/ (the Wellcome Trust Sanger Institute).

Serotype	Strain	Plasmid Name	Size (bp)	Source
Typhimurium LT2	SGSC1412	pSLT	93939	NCBI: gb|AE006471.1|
Choleraesuis	OU7519	pOU7519	127212	NCBI: gb|EU219534.1|
	SC-B67	pSCV50	49558	NCBI: gb|AY509003.1|
Dublin	CT_02021853	pCT02021853_74	74551	NCBI: gb|CP001143.1|
		pOU1115	74589	NCBI: gb|DQ115388.2|
	OU7025	pOU1113	80156	NCBI: gb|AY517905.1|
Enteritidis	P125109		59372	Sanger
Gallinarum	287/91		88350	Sanger

Table 8. Salmonella virulence plasmids in database.

Annotation

Annotation has been performed using the Computational Biology at Nano+Bio-Center (http://nbc3.biologie.uni-kl.de/). Firstly, we have uploaded the pSS44 sequence in a FastA format file in the Glimmer tool and we obtained the predictive coding sequences (CDSs) with the corresponding protein sequences. Therefore, we have used the Best Blast Annotation program to search for the similarity between the protein sequences and the protein already characterised (BlastP).

Lambs experimental infection

One to two month-old Sarda lambs with no cultural or serological evidence of *Salmonella* infection were used. Lambs were infected subcutaneously with about 10^9 CFU of each strain. Inocula were obtained by growing strains statically at 37°C over night. After one, two and three weeks, rectal temperature, deambulation problems and inflammation signs were examined. Additionally, blood samples were taken every week to evaluate immunisation induction via an indirect enzyme-linked immunosorbent assay (ELISA) using the serotype Abortusovis outer membrane protein (OMP) crude preparation described elsewhere (Uzzau, *et al.*, 2005). Animal experiments were performed at the laboratories of Istituto Zooprofilattico Sperimentale della Sardegna (Sassari).

Infection of THP-1 cells

THP1-Blue cells (InvivoGen) were used for infection in vitro (kindly supplied by Daniele Dessi' from Department of Biomedical Sciences, University of Sassari). The activation of NF-κB has been valued by a colorimetric assay that determines secreted alkaline phosphatase (SEAP) activity (the SEAP gene is plasmid-encoded and it is under the control of a promoter inducible by several transcription factors such as NF-κB).

Cells were grown in RPMI 1640, 10% fetal bovine serum with 2 mM L-glutamine and Pen-Strep (Gibco) adjusted to contain 1.5 g/L sodium bicarbonate, 4.5 g/L glucose, 10 mM HEPES, 1.0 mM sodium pyruvate. After some passages, Zeocin 200 μg/mL (InvivoGen) has been added to the growth medium to select cells with the plasmid.

When the cells reached an optimal amount, they were activated with phorbol myristate acetate (PMA, Sigma) 50 ng/mL, that induced their differentiation in adherent macrophage-like cells. 500.000 cells/mL were seeded in a 96-well microtiter plate (200.000 cells per well). After 24h, 72h and all following days (till at 7^{th} days of differentiation) we changed growth medium (without PMA). The last medium was without Pen-Strep to avoid killing of bacteria.

Selected strains were growth in LB broth over night. The following day, bacteria were washed in PBS buffer and resuspended in PBS int 10^8 cells/mL. Infections were performed with 10 μL of suspension, with a multiplicity of infection (MOI) 10:1. Also, LPS (positive control) was inoculated in scalar concentrations (from 10 to 10^{-4} μg/mL).

After 105 minutes, we changed medium and added Gentamicin (Sigma) 100 μg/mL for 1 hour to kill all bacteria external to the cells. Then, we changed medium with Gentamicin 10 μg/mL and 24 hours to infection, surnatant was taken.

For each sample, 2.5 μL of surnatant were added to 50 μL of p-Nitrophenyl phosphate medium (Sigma) and after 1 hour we read absorbance at 405 nm.

Chapter 4

RESULTS

Structural genomics

Genome sequencing of SS44

The whole-genome sequence for *S. enterica* serotype Abortusovis strain SS44 has been generated using high-throughput 454 technology. The analysed genome is composed by a chromosome and a virulence plasmid (with estimate lengths of about 4.8 Mb and 65 kb respectively). Joining this technology with traditional sequencing, we produced a primeval map of the plasmid and different comparisons of chromosomal regions among different serotypes.

454 sequencing

After three 454 sequencer runs, we reached nearly half million of short sequences (reads) for almost 52 Mb (Table 9). We estimate an 11-fold coverage with respect to the size of *S. enterica* serotype Typhimurium strain LT2 (about 4.8 Mb). The pyrosequencing software assembled the reads into contigs (Table 10). The software employs the sequence overlap of the reads to reach a consensus that will be used to write the contig sequence. Unfortunately, the software cannot assemble all reads, especially repeated regions; then there were several contigs.

Together with the last run, paired-ends were been performed. This analysis was used to obtain a set of scaffolds. Each scaffold is composed by one up to several contigs separated by gaps. There are two kinds of gaps: between each contigs of the same scaffold (intra-scaffold gaps) and between scaffolds (inter-scaffold gaps). The software is able to hypothesize the intra-scaffold gaps dimensions whereas does not return any information regarding inter-scaffold gaps.

Run	1	2	3	1+2+3
Reads #	118932	320229	39424*	478585
Length min (nt)	37	38	35	35
Length max (nt)	176	166	342	166
Average (nt)	99,5	101,7	191	108,4
Total nt	11853610	32481968	7547550	51883128

Table 9. Reads. *The third run was performed together with paired-ends then there are less reads but more long because has been used the new pyrosequencing machine.

	Contigs	Scaffolds
Sequences #	702 (567*)	124
Length min (nt)	101	679
Length max (nt)	94498	190166
Average (nt)	6357	36688
Total nt	4462413	4549301

Table 10. Contigs and scaffolds. *Only these contigs are included in scaffolds.

Gap closure

Following automated sequencing, research activities were focused on gap closing. In deed, software did not assemble all reads in contigs and this task, therefore, has been accomplished manually. Firstly, intra-scaffold gaps have been filled, since the terminal sequences of each gaps are available for 454 sequencing. Initially, there were 124 scaffolds with 443 internal gaps (ranging from 20 to 1576 estimated nucleotides). To fulfill gap closure, all reads complementary to the end of the contigs were sought. Often, several reads protruded into the gap, allowing to narrow it or even close it (Figure 3). After this step, 100 intra-scaffold gaps remained to be filled (ranging from 20 to 1498 estimated nucleotides). The same approach was performed on inter-scaffold gaps. This has allowed to link different scaffolds and contigs in super-scaffolds. All together, 90 scaffolds and 17 super-scaffolds (ranging from 29181 to 232988) were assembled (Table 11).

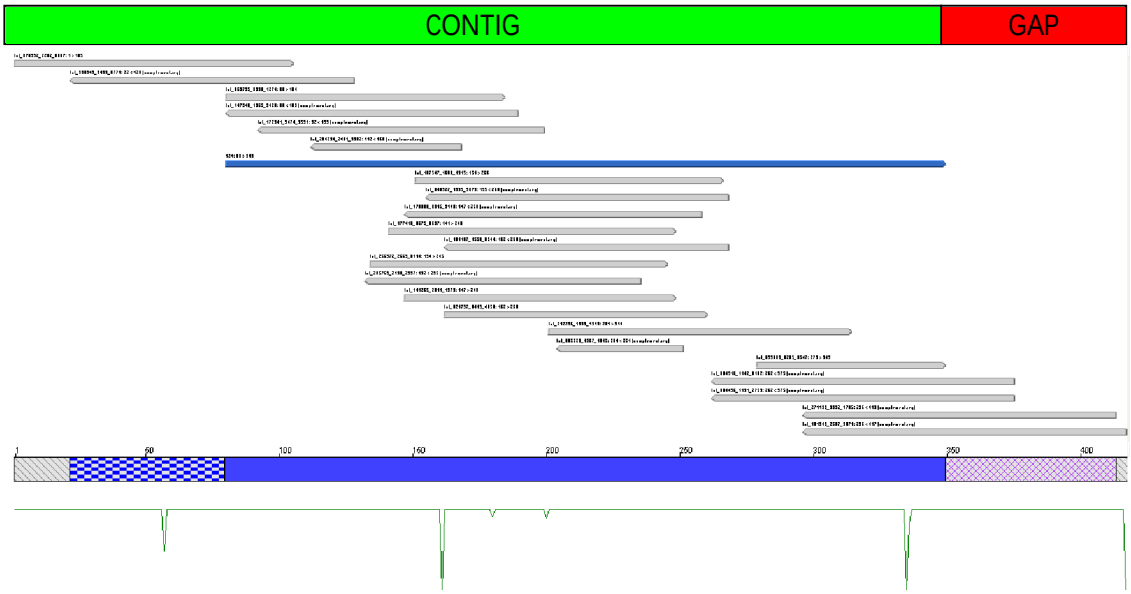

Figure 3. Gap closure. Figure shows an example of reads alignment on a query sequence; in this case the sequence (the upper blue arrow) is the terminus of a contig (green bar). The reads which had a similarity with this sequence are showed as grey arrows. It can be noted that some reads extended the contig into the gap (on the right of the contig sequence). They have been used to narrow the gap (red bar).

	Scaffolds	Super-scaffolds
Sequences #	90	17
Length min (nt)	679	29181
Length max (nt)	170298	232988
Average (nt)	29543	108858
Total nt	2658838	1850578

Table 11. Scaffolds and super-scaffolds after gap closure. The sum is smaller of the sum of the old assembled scaffolds because some scaffolds are included inside the gaps.

Coverage analysis

Using reads coverage (how many reads were complementary with a specific region) it is possible to speculate on copy number (CN) of a specific locus (Table 12). Among repeated regions, about hundred copies of IS*1414* may be extimated, even though *astA* gene (internal to the insertion sequence) has a higher CN. This might be due to the fact that some IS1414 are not complete. The presence of 7 copies of *rrl* genes (as noted for other *Salmonellae*) (S. L. Liu*, et al.*, 1993) were confirmed, as four copies of IS*200* (Schiaffino*, et al.*, 1996), of each one in the virulence plasmid (see below). In addition, about 13 copies of IS*21* were estimated.

Locus	bp locus	bp reads	coverage	copy number
		Repeated regions		
IS*1414*	1209	1264182	1045.64	95.06
astA	117	158876	1357.91	123.45
IS*200*	507	25287	49.87	4.53
IS*21*	798	115731	145.02	13.18
*rrl**	2612	196689	75.30	6.85
		Plasmidic regions		
spvR	894	10813	12.10	1.10
spvA	768	8626	11.23	1.02
spvB	1782	19820	11.12	1.01
spvC	726	8460	11.65	1.06
spvD	651	6615	10.16	0.92

Table 12. Coverage of repeated and non repeated regions. Bp reads is the sum of the lengths of all reads that composed a region. Coverage is the ratio of bp locus and bp reads. Copy number is calculated as the ratio of coverage in the sequence and the medium coverage (11x). The copy number medium of the *spv* genes is 1.02. **rrl* is actually a contig who contains a I-*Ceu*I restriction site and then represent all copies of *rrl* genes.

Copy number of *spv* operon genes, according to that of the virulence plasmid, appears to be equal at one per chromosome. Considering these results, strain SS44 genome showed a length of about 4.51 Mb (scaffolds and super-scaffolds sum) plus about 150 kb (a hundred copies of IS*1414*). This is a total of at least 4.7 Mb.

Virulence plasmid sequence and physical map

The second step of the work was the assembly of the virulence plasmid (named pSS44). Firstly, all contigs complementary to the virulence plasmid of *S. enterica* ser. Typhimurium strain LT2 (pSLT), chosen as reference, were searched (Table 8). This method allowed selecting three scaffolds with an overall length of 45 kb (Scaffold #116, #111 and #94). Then, contigs without complementarities with reference's chromosome were sought (SAO-specific) allowing selection of scaffold #016 (Table 7). Rather the most of scaffold #016 shows about 90 % of similarity to *E. coli* plasmid p53638_226. This region was about 11 kb in size and further analysis showed short sequences (about 3 kbp) with homologies (70 %) with a *S. enterica* serotype Dublin virulence plasmid, as described below. To confirm the presence of the scaffold #016 in the plasmid, a PCR amplification of a 800 bp sequence within the scaffold was performed. Using strains SS44 and SU40 (a SS44 cured) as template, amplification has been achieved only in SS44. Therefore, scaffold #016 belongs to pSS44 virulence plasmid (Figure 7).

Aligning of scaffolds and gap closure was obtained as above with molecular biology methods (PCR). Plasmid map was obtained by restriction fragment length profiling method of pSS44 with different enzymes (Figure 4).

Overlay of restriction analysis data, gap closure, and scaffolds alignment pattern, allowed a primeval map of pSS44. This map originates from the assembly of four scaffolds (scaffolds #016, #116, #111 and #94) and as much inter-scaffold gaps (IN1, IN2, IN3 and IN4). IN3 gap was closed *in silico*, then a set of primers were designed pointing outwards each scaffolds to amplify the sequences of remaining gaps. Besides, primers for intra-scaffold gaps (IT1, IT4 and IT7) were been designed. Amplicons were sequenced via Sanger technology. In this way has been possible closing all gaps but the inter-scaffold two gaps (IN1 and IN4) flanking the scaffold #016 (because there was not amplification). To date, the two gaps were not amplified yet. However, a provisional map of pSS44 has been obtained (Figure 5).

The plasmid sequence shows a GC medium content of 49.7 % but a region with a higher value has been identified. This sequence corresponds to a putative operon harbouring genes with homology with EAL (putative cyclic diguanylate phosphodiesterase), the *rds* resolvase, and the genes encoding the plasmid maintenance proteins *ccdA* and *ccdB*. It is noteworthy that scaffolds #016 owns a GC bias ((G-C)/(G+C)) different from the other regions.

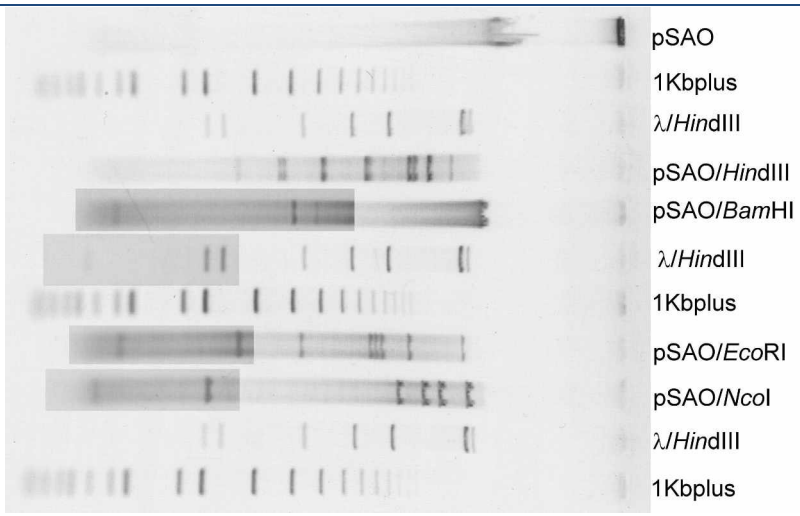

Figure 4. Restriction pattern. 1st lane: undigested pSS44 DNA; 2nd, 7th and last lane: 1 Kb plus marker; 3rd, 6th and 10th lane: λ phage DNA digested with *Hin*dIII; 4th lane: pSS44 digested with *Hin*dIII; 5th lane: pSS44 digested with *Bam*HI; 8th lane: pSS44 digested with *Eco*RI; 9th lane: pSS44 digested with *Nco*I. This image derives from two images acquired with different exposition to show clearly all bands.

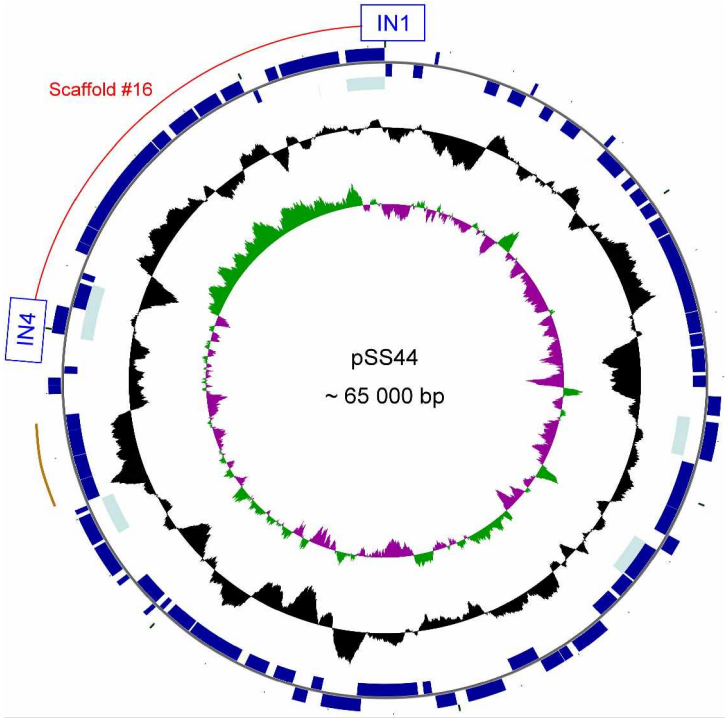

Figure 5. Graphical representation of provisional pSS44 sequence. The scaffold #016 is red-lined and gaps inter-scaffold IN1 and IN4 are indicated but the location of the scaffold in the plasmid could be different. The putative operon is brown-lined. Outer circle shows CDSs as blue arcs on forward and reverse

strand. Arcs light blue are IS*1414*. In the central circle is represented G+C content and the inner one displays GC bias; purple indicates value > 1; green < 1.

Annotation of pSS44 plasmid sequence

The provisional sequence of the plasmid pSS44 has been annotated. 90 putative coding sequences (CDSs) have been automatically identified, then we added manually other four genes, comparing with genome already annotated, reaching 94 total CDSs named pSS44_01 to pSS44_94 (Table 13).

The virulence plasmid harbours the *spv* operon (as the all *Salmonella* virulence plasmids described), an uncomplete set of genes involved in conjugation, and a set of fimbrial genes. Moreover, there are probably six IS*1414* insertions (with some non functional *tnpA* genes), one IS*200* and one IS*21*. Also, pSS44 encodes for the chaperon protein ClpE and ClpF that are homologue only with *E. coli* strains and furthermore, the plasmid harbours fimbrial genes with homology (80 %) to FaeI K88 fimbrial protein A and CshB porin of *S. enterica* serotype Schwarzengrund.

Functional Class	number
Hypothetical	29
Transposase	21
Fimbrial	10
Conjugative	11
Virulence	7
Replication	10
Enzime	5
Regulator	1
Resistance	0
Total	94

Table 13. Functional classes of the gene encoded by pSS44.

Comparison between pSS44 and other SAO plasmids

To valuate the stability of the genomic structure in virulence plasmids of serotype Abortusovis, a digestion with *Hin*dIII of different plasmids of SAO was performed. Five restriction patterns (R1 to R5) and some fragments common to all strains, corresponding to about 2500, 3700 and 5000 bp in length, were identified (Figure 6). According to the pSS44 annotated sequence, they would correspond to pSS44_03-04, pSS44_49-53 (a part of *spv* operon) and pSS44_05-12 (some of which homologue to *yacAB* and *yagA* genes), respectively.

Different geographical origin strains are associated with a different restriction pattern with exception of those IS*1414*$^{-}$ (that hold the same digestion pattern R5) and SS44, which share the 3253 (English strain) pattern.

These evidences were validate performing a PCR amplification with primers for *yacAB* and pSS44_90 (located in scaffold #016) in all SAO strains (Figure 7). All strains harbour *yacAB* region, but only IS*1414*$^{+}$ SAOs have the pSS44_90.

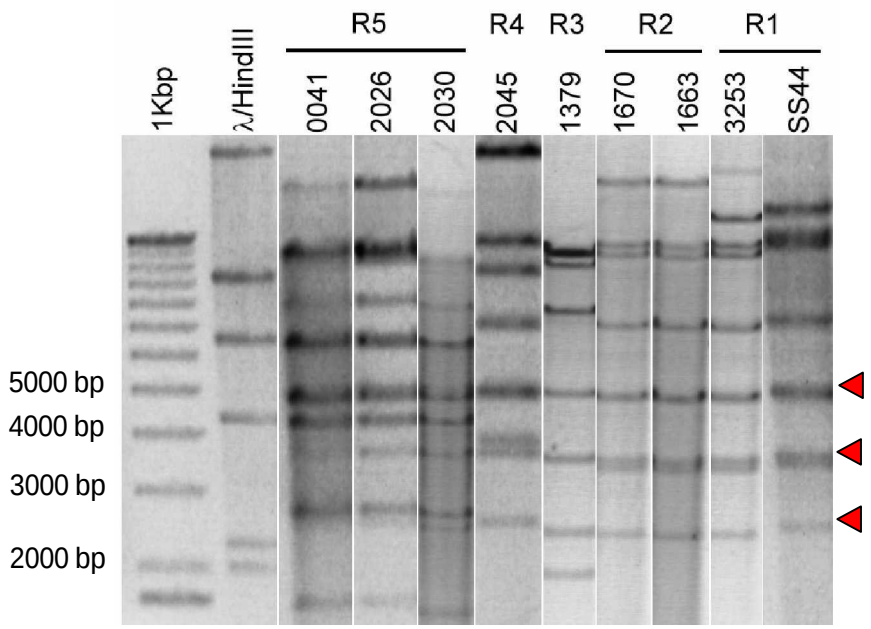

Figure 6. Restriction patterns of virulence plasmid of Abortusovis strains. The first lanes are the markers 1Kb plus and λ DNA/*Hin*dIII. The lanes are ordered according to the five restriction patterns (R5 to R1). The red arrows indicate all common fragments.

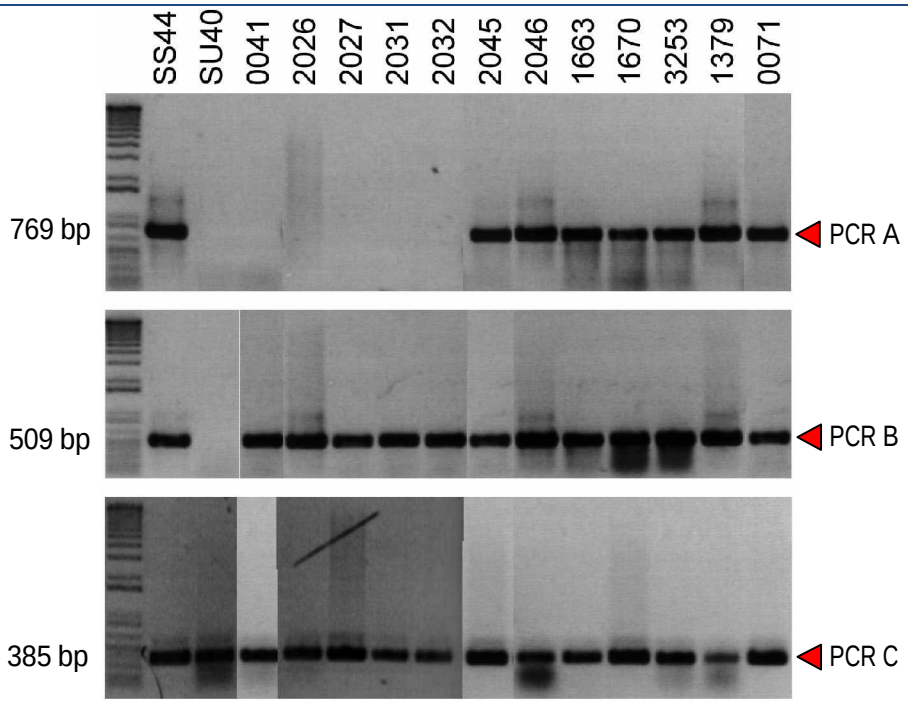

Figure 7. The figure shows the results of the PCRs. PCR A: *pSAO0090*; PCR B: *yacAB*; PCR C: *avrA*, a chromosomal gene chosen as positive control. SU40 strain does not carry the plasmid, hence only PCR C is positive.

Comparison between pSS44 and other *S. enterica ssp.* I serotypes

To date, the sequences of several virulence plasmids are published (Table 8). The comparison between pSS44 and the *S. enterica* ser. Typhimurium strain LT2 virulence plasmid (pSLT) shows close homologue regions such as the *spv* operon and locus *tra* (Figure 9 and Table 14). A comparison with other virulence plasmids available demonstrated that scaffold #016 does not show any similarity but a 73% homology with pCT02021853_74 plasmid (*S. enterica* serotype Dublin) (data not showed).

Rather, blast analysis of the scaffold #016, retrieves a short homology with the chromosome of *S. enterica* serotype Schwarzengrund (see above).

Functional (gene ontology) comparison between pSS44, the reference pSLT, and the virulence plasmids of *S. enterica* serotypes Choleraesuis strain SC-B67 (pSCV50) was performed (Figure 8). The main differences reside in the percentage of transposases content (higher in pSS44, because of the presence of IS1414) and conjugation related genes (pSLT is conjugative as opposite to pSS44).

Gene	Size (bp)	% Homology Nt	% Homology Aa
spvR	894	99	99
spvA	767	99	99
spvB	1782	99	99
spvC	726	99	99
spvD	651	99	98

Table 14. Plasmidic virulence genes. In this table the gene length, percentage of homology in nucleotides and aminoacids are indicated.

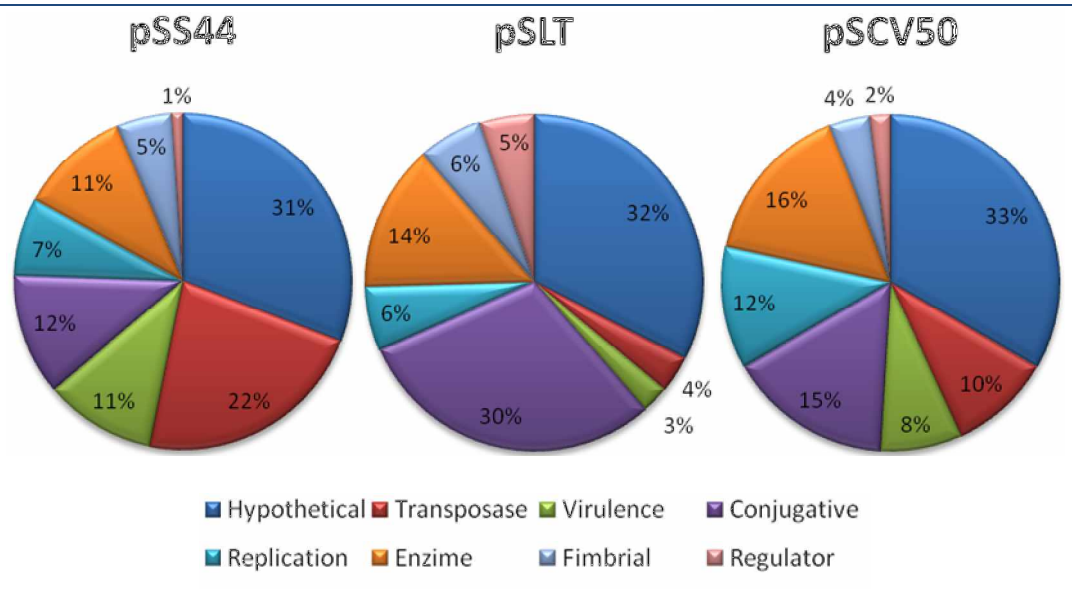

Figure 8. Assortment of the genes according to the function in three virulence plasmids.

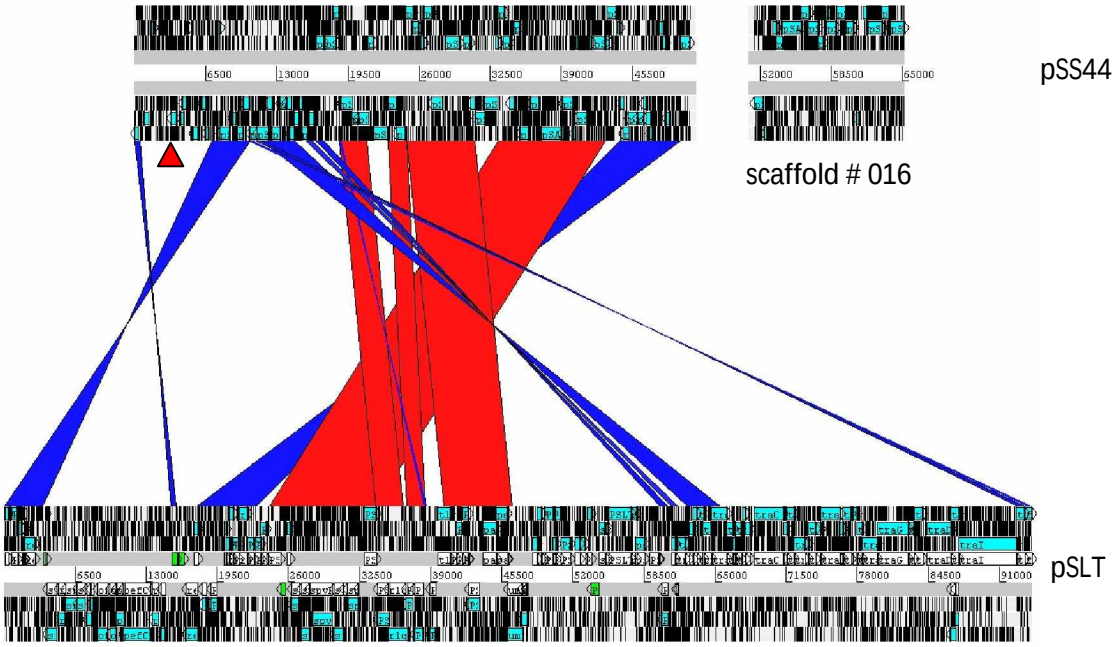

Figure 9. Comparison between pSS44 and pSLT. This drawing represents the two sequences and the regions with similarity are matched by strips red or blue (direct or reverse homology, respectively). The scaffold #016 sequence is shown separated by other scaffolds. pSS44 shows strong homology with pSLT, except the scaffold #016 and *yacAB* and *yagA* loci (red arrow).

Pulse field gel electrophoresis (PFGE) of SAO strain

Macro-restriction of different Abortusovis strains was performed using the PFGE. SS44 shows a unique pulsotype P1. All isolates IS*1414*$^{-}$ share the same pattern (P3), which is different from Iranian IS*1414*$^{+}$ strain (P2). Sardinian and English strains show the same pattern (P4). French and Albanian strains, and SS44 showed unique pulsotypes (P5, P6 and P1, respectively).

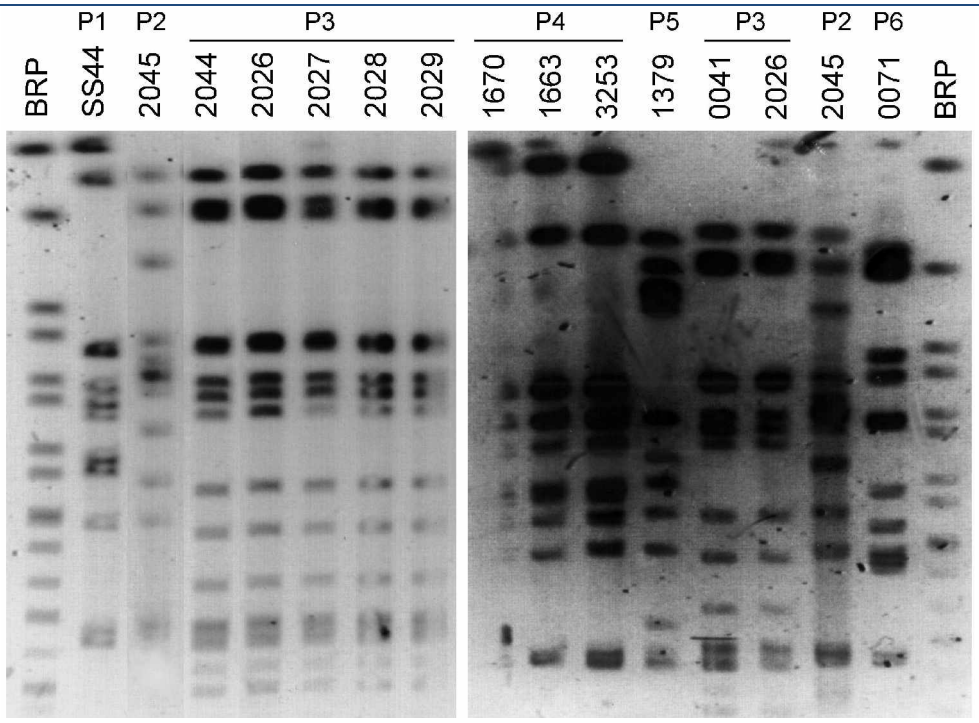

Figure 10. This figure shows two PFGE gels. BRP means the *S. enterica* ser. Braenderup strain used as reference marker.

Salmonella pathogenicity islands of SS44

The SPIs 1 to 5 have been partially assembled and annotated. Then, SS44 SPI sequences have been compared to others harboured by different serotypes (Table 7).

SPI-1

The scaffolds #020 (24149 bp) and #105 (23164 bp), separated by IN-SPI-1 gap, encompass the SS44 SPI-1. The gap is localised between *iagB* and *sptP* genes (Figure

11). Because the two genes are contiguous in all *Salmonellae* genomes analysed, we performed a PCR to amplify the gap to sequence it. The PCR showed that the two scaffolds are not contiguous in SS44 (Figure 12).

The locus *sptP-iagB* is a hot spot of IS*1414* insertion; in fact, all IS*1414*$^{+}$ SAO strains analysed carry the insertion in this point (Figure 13).

The pathogenicity island 1 of SS44 strain has been compared with SPI-1 of other strains. SPI-1 is conserved throughout the whole genus *Salmonella*. However, we report the lack of *avrA* gene in *S. enterica* serotypes Choleraesuis (SCS), Typhi (STY), Paratyphi A (SPA)(Bacciu*, et al.*) and *S. enterica ssp. arizonae* serotype *62:z4,z23:--* (SAR). Also, SPI-1 in *S. bongori* (SB) has a lower homology with the other SPIs, missing most of *sptP* gene. In SS44, SptP protein is 438 aa long, whereas STM and *S. enterica* serotype Dublin (SD) SptP are 543 aa and 535 aa long, respectively. This discrepancy is due to single base switch leading to a stop codon in the 5' end of the ORF. In SS44, SptP is expected to have a Met start corresponding to Met107 of STM SptP.

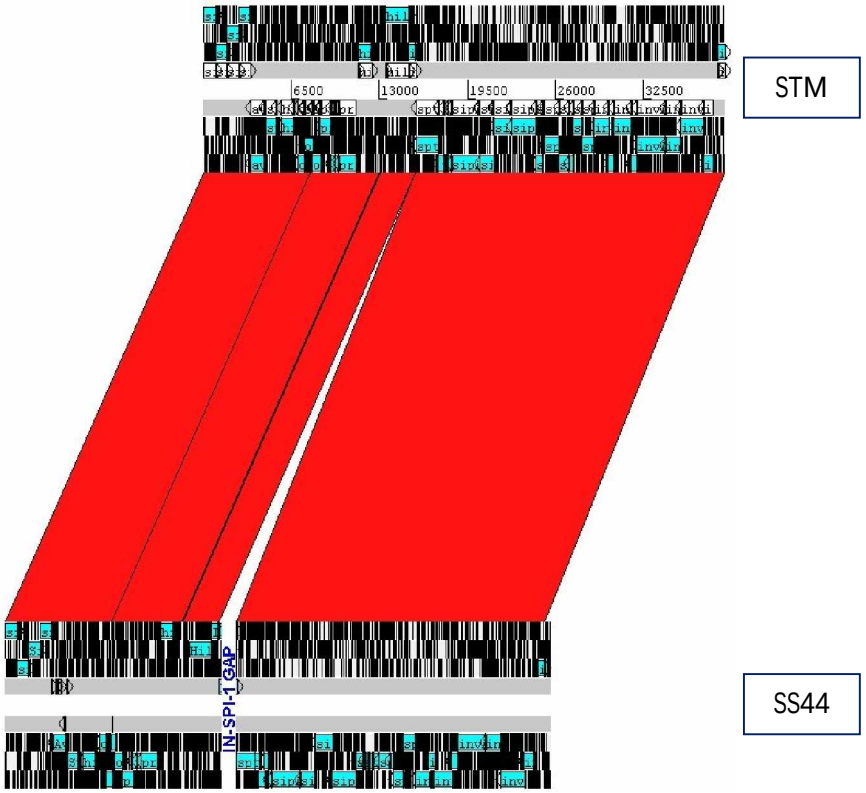

Figure 11. Comparison between SS44 and STM SPI-1. We indicated the gap in the SS44 sequence, so far, we do not know the length, it could contain several scaffolds inside. The comparison show high level of homology between two sequences.

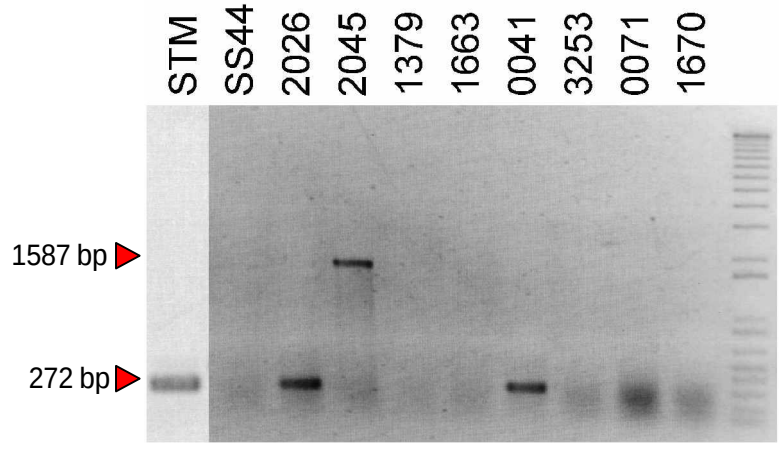

Figure 12. PCR IN-SPI-1 result. All IS*1414*$^{-}$ strains (STM, 2026 and 0041) amplified an about 300 bp fragment (272 bp in STM, obtained by the sequence); all IS*1414*$^{+}$ strains do not amplify, with the exception of the 2045 strain, probably because the *iagB* and *sptP* are very distant.

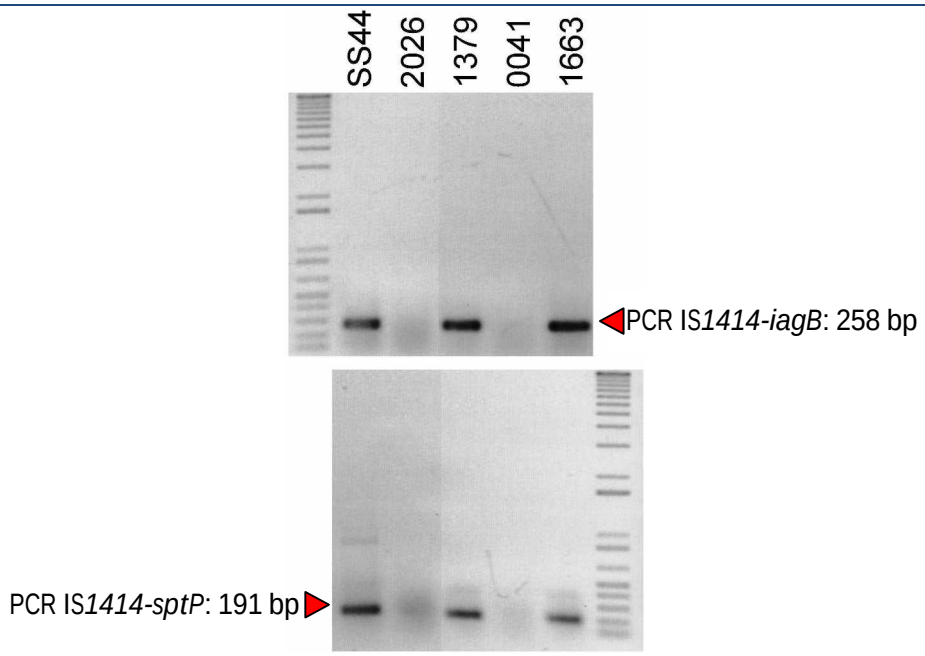

Figure 13. PCR IS*1414-iagB* and IS*1414-sptP*. The first PCR was accomplished with 08-29 and 02-62 primers; the second one with 08-30 and 02-63 primers. These PCR amplifications demonstrated the IS*1414* insertion in the locus *iagB-sptP* for SS44, 1379 and 1663 strains.

SPI-2

We found SS44 SPI-2 in a part of the scaffolds #037 and #114. The sequence analysed harbour four gaps intra-scaffold and one gap inter-scaffold (IN-SPI-2). Analysis of the island showed some differences between serotypes. The comparison between STM and SS44 has shown only a lack of a region (that corresponds to STM orf408) in SS44. In fact the orf408 in SAO is long only 339 bp rather than 1224 bp in STM. This open read frame (ORF) is long about 1224 bp in all serotypes analysed except SAR. SAR lack not only the orf408 but about 3000 bp at 5'. On the other end, SB

possesses this ORF but have also an insertion between orf319 and orf70 in SAO. Again, the SPI-2 of SB undergoes a wide deletion (about 22 kbp) of the 3' portion.

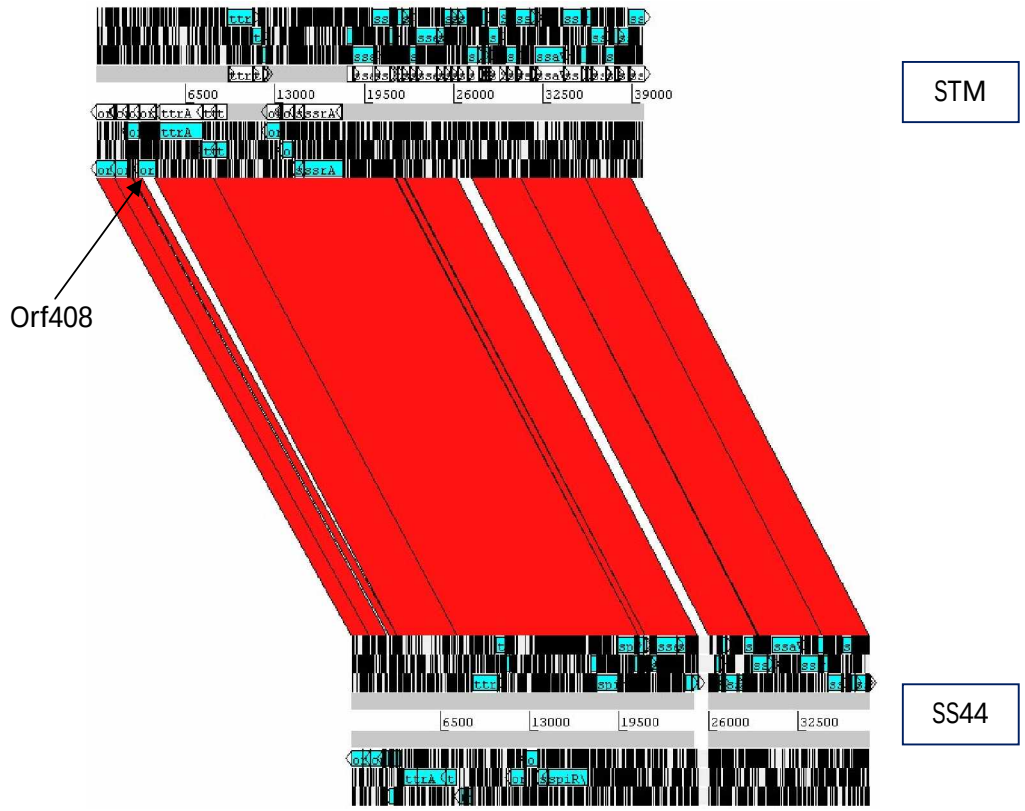

Figure 14. Comparison between SS44 and STM SPI-2. It is shown the gap IN-SPI-2 in the SS44 sequence, we supposed 798 bp in length, but it is not sequenced yet. There are differences in regions corresponding to the gaps and orf408 (arrow)(Thomson *et al.*).

SPI-3

The SS44 SPI-3 has been composed with the scaffolds #003 and #103. An inter-scaffold (IN-SPI-3) and an intra-scaffold gap were found. The island has revealed extensive variation in its structure with respect to STM. In fact, SS44 carries two deletions: the first one between *selC* and *rhuM* (in the scaffold #003) and the second one be-

tween *misL* and *slsA* (among the two scaffolds). It is noteworthy that scaffold #103 harbours several genes upstream *slsA* that do not belong to the SPI-3. Therefore this island is broken in two parts and a previous sequencing of the *misL* region showed the presence of IS*1414* (Figure 15).

SHA, SIN and SCS lack *sugR* and *rhuM* as SS44. SAR owns SARI03891 gene inserted between *yicH* and *yicI*. This is a putative protein and don't have a similarity with other gene. Also, SAR lacks *rmbA* and *misL*. Finally, SB SPI-3 lacks *sugR*, *rhuM* and *misL*.

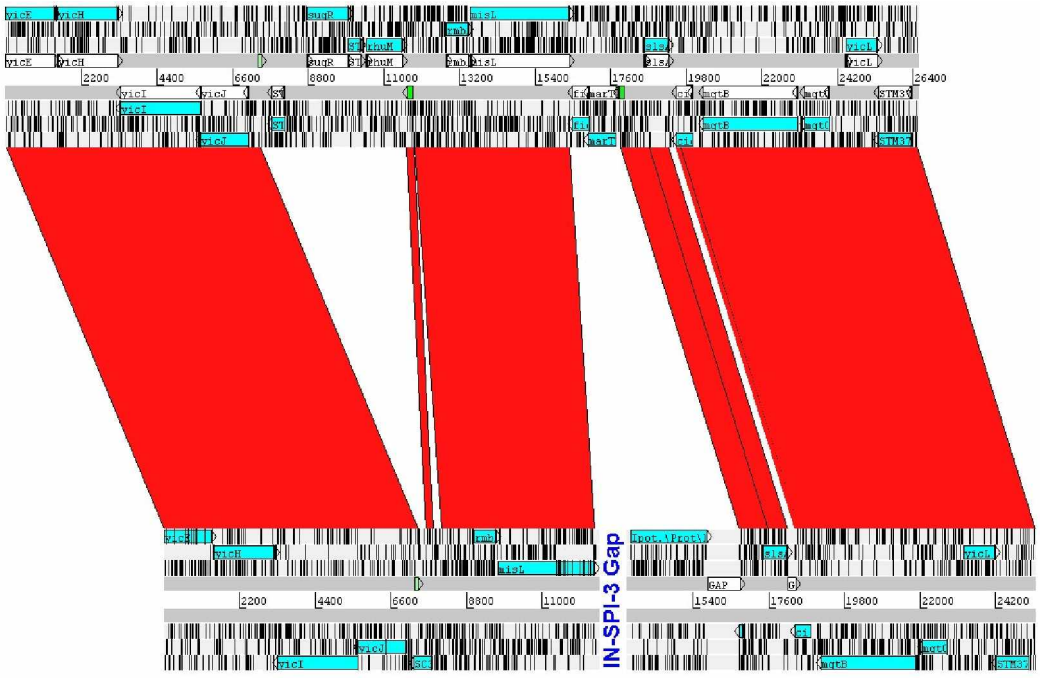

Figure 15. Comparison between SS44 and STM SPI-3. Two intra-scaffold and the IN-SPI-3 gaps are displayed in the SS44 sequence; downstream to *misL* there is an IS*1414* (not shown). SS44 lacks about 5 kbp (encompassing *sugR* and *rhuM*) and *marT* locus.

SPI-4

The SS44 SPI-4 has been identified in the scaffolds #028 and #076, separated by IN-SPI-4 and other six intra-scaffold gaps. The island is quite similar in all serotypes analysed (Figure 16). The *siiABCDEF* operon is ever present but SAR presents less homology among the other.

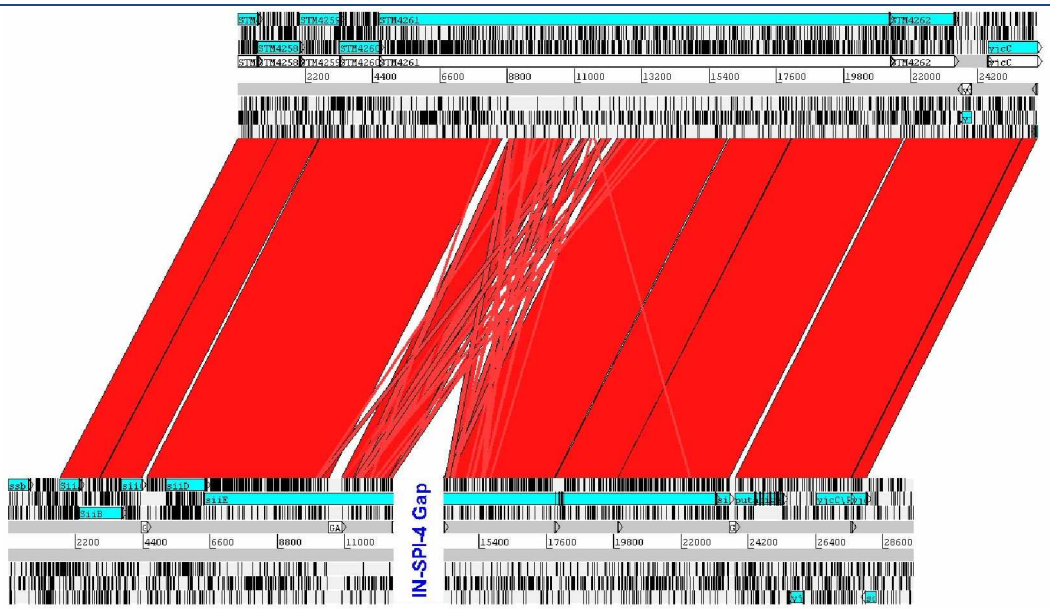

Figure 16. Comparison between SS44 and STM SPI-4.The *siiE* gene is truncated by the inter-scaffold gap. There is high level of homology (the only breaks corresponding to the gaps). The tangle in the central part of is due to a repeated region inside the *siiE* gene (which has hampered assembly of two scaffolds).

SPI-5

The SS44 SPI-5 has been identified in the scaffold #087 (which owns an intra-scaffold gap). The *pipA* gene encoded in SS44 is shorter than STM (327 bp rather than

681 bp) (Figure 17). Also, SB and SAR have an insertion (about 8 kb) downstream *sopB* gene with respect to SS44.

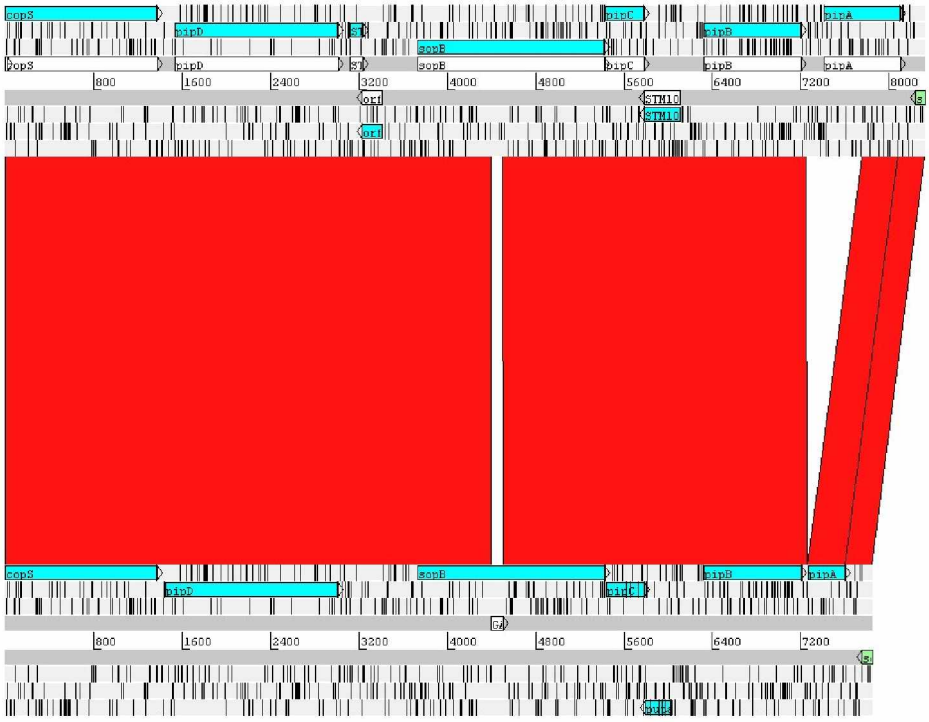

Figure 17. Comparison between SS44 and STM SPI-5.The deletion in *pipA* locus is the only difference between the sequences.

Other virulence genes

SS44 encodes other genes involved in virulence which are not encoded by SPIs: sopA (scaffold #070), sopD (scaffold #017), sopD2 (scaffold #086) and sopE2 (scaffold #006) (Table 15). The *sopA* and *sopE2* sequences include a gap intra-scaffold (238 bp and 299 bp, respectively).

Gene	Size (bp)	% Homology Nt	% Homology Aa
sopA	incomplete	99	98
sopD	947	99	98
sopD2	955	97	96
sopE2	incomplete	100	99

Table 15. Other virulence genes. Homology is calculated with respect to STM. Percentage could be inaccurate in incomplete sequences.

Functional analysis

Taking into account previous infection experiments suggesting a possible involvement of pSS44 in inflammation (Bacciu, Marogna and Uzzau, unpublished results), two series of experiments (*in vivo* ad *in vitro*) were performed to clarify in this respect, a possible role of pSS44 encoded genes. To test the hypothesis, recombinant SU40 strains harbouring various fragments were injected to lambs subcutaneously. Animals were monitored for signs of local inflammation. Furthermore, NFκB activation was assessed *in vitro* by stimulating THP-1 cells with the same mutant strains.

Infection *in vivo*

Two different experiments *in vivo* have been performed: the first with nine strains and the second with the strains with a more significative effect. All strains have showed an adequate immunisation. In fact, ELISA analysis detected antibody production versus SS44 antigens, as compared to the PBS inoculum in lamb (Figure 18B). Lambs infected with SS44 and SU40 strains both exhibited a swelling, as also 4323, 4324, 4325 and 4327 strains infection. Conversely, infection with 4326 and 4327 strains induced less inflammatory phenomena. It is noteworthy that 4320 strain pro-

duced only slight inflammatory signs after the first week but the animal did not show any visible sign after two weeks (Figure 18A).

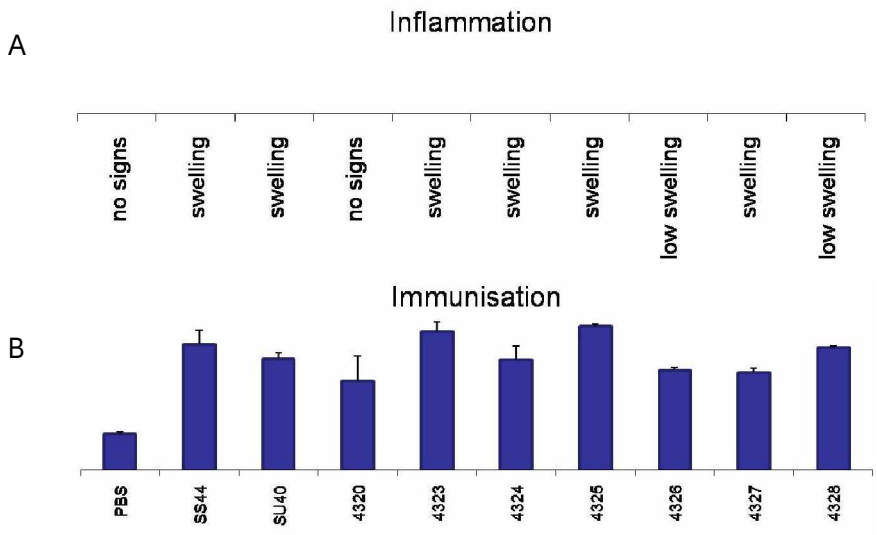

Figure 18. First experiment: inflammation signs and immunisation levels. A: Inflammation signs are defined after third week to infection by veterinary examination. B: Immunisation levels are calculated via absorbance detection using ELISA method.

The second experiment has been performed with infections in two different animals per strain, and confirm previous evidence. In fact, 4320 strain showed again no clinical symptoms after three week of infection, while SS44, SU40, and 4323 strains induced serious swelling (Figure 19B). All strains induced raise of antibody titers, as confirmed by ELISA (Figure 19A).

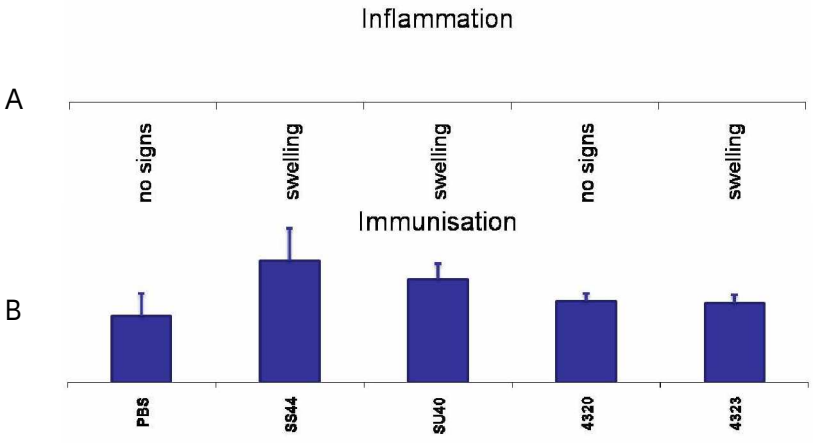

Figure 19. Second experiment: inflammation signs and immunisation levels. A: Inflammation signs. B: Immunisation levels.

Infection *in vitro* (THP-1 cells)

We carry out a set of experiments on cell cultures in order to evaluate if a lower inflammatory induction by 4320 strain could be dependent on reduced NF-κB expression. Activation of NF-κB pathway has been evalued using THP1-Blue cells infection model. As described in Methods, activation of NF-κB can be measured by colorimetric assay.

Four experiments have been done using SS44, SU40 and 4320 strains. In each experiment, THP-1 cells were inoculated with lipopolysaccharide (LPS) as positive control and PBS as negative control (Figure 20). The experiments showed discordant data; therefore a reduction of NF-κB activation by 4320 strain was not assessed.

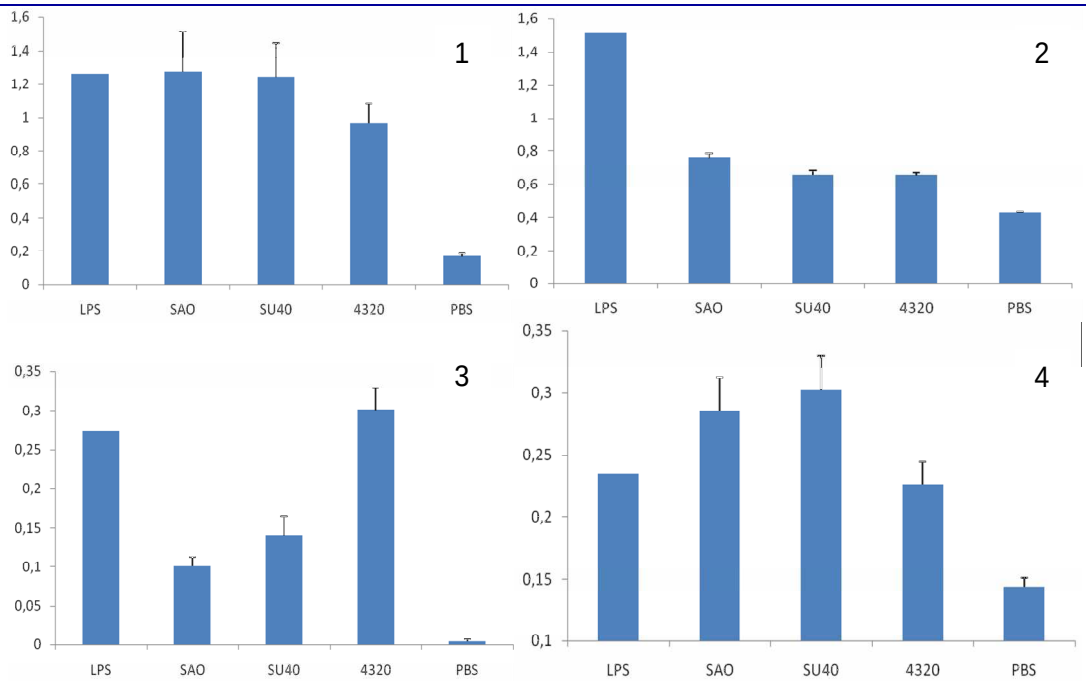

Figure 20. Activation of NF-κB THP-1 cells. For each experiment, LPS (positive control) and three SAO strains are tested.

Molecular characterisation of 4320 cloned fragment

Since 4320 strain has showed to reduce local (sub-cutaneous) inflammation, the content of pDB108 cloned fragment was further characterised. pDB108 was partially sequenced to identify the right *Hin*dIII fragment in the pSS44 assembled sequence. The fragment is 7326 bp in length and harbours ten CDSs (Figure 21). An IS*1414* including *tnpA* and *astA* genes has been identified; it is noteworthy that this *astA* is the only one functional in the plasmid; in fact, the other IS*1414* encompass *astA* ORF with frameshift mutations. Furthermore, an IS*200* insertion sequence was recognised. Other CDSs included: an integrase-like protein (Y4LS), *mig*-5 (a gene encoding a

putative β-class carbonic anhydrase probably involved in virulence), *tlpA* (a temperature-sensing protein with a possible role in the *Salmonella* pathogenicity), *yacC* (a putative periplasmic protein) and other putative protein coding sequences.

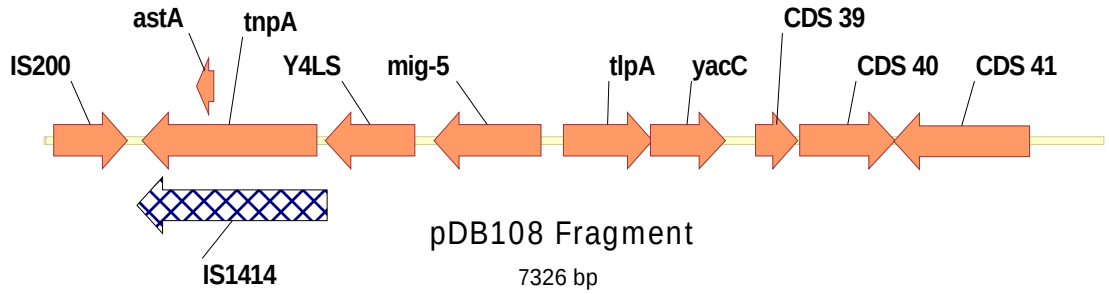

Figure 21. Map of the fragment cloned in pDB108. We reported all annotated CDSs.

Chapter 5

DISCUSSION

The possibility to investigate the molecular basis of pathogenicity and host-adaptation by microbial pathogen is greatly enhanced by the availability of sequencing data and, as a consequence, the feasibility to associate loci to functional biochemical pathways, virulence determinants, and clinical traits.

Since the first bacterial whole genome sequence has been released, a large group of genomes, increasing rapidly every year, have been available (Fleischmann*, et al.*, 1995). This progress has been certainly facilitated by an impressive improvement of high-throughput sequencing technologies, thanks to the wide sequence studies on human genetics and genomics. Bacterial genomes, considering the relatively small lengths and complexity, have become easily approached by the use of such new sequencing technologies (Medini *et al.*, 2008).

As the microbial world entered the Post-Genomic Era, it is now possible to pose several questions on overall association between bacterial genotype and phenotype, so far unanswered.

The main reason to tackle the whole sequencing of *Salmonella enterica* ser. Abortusovis was to create a genetic database to be exploited by a wide array of investigations on bacterial pathogenesis and biology on *Salmonella* infection.

Expected results from such studies will have a great impact on the knowledge of molecular basis of virulence and host-adaptation, also in other relevant pathogens, including human-restricted serotype Typhi.

S. enterica serotype Abortusovis strain SS44 has been sequenced by making use of 454 sequencing technology (pyrosequencing).The 454 sequencing combines cheapness with high-throughput. In the last years several genome drafts sequenced using 454 have been published (Goldberg *et al.*, 2006; Stiens *et al.*, 2008; Wicker *et al.*, 2006). However, it is not possible to sequence a whole genome *de novo*, using solely pyrosequencing. Indeed, this method provides high coverage but short sequences and such feature makes difficult the assembly of repeated regions. Therefore, the present work required a combination of pyrosequencing and classical sequencing methods as Sanger technology.

One of the main results reported in this thesis project was the identification and the assembling (although incomplete) of the scaffolds belonging to the virulence plasmid pSS44, using different bioinformatics and molecular biology approaches. Besides, coverage analysis allowed to estimate the copy number of IS*1414*. An IS that, as confirmed in this study, may play an extremely important role in genome plasticity and pathogenicity of serotype Abortusovis.

Furthermore, comparative analysis between SS44 sequences with other *Salmonellae* led to identify major differences in SPI-1, SPI-3, and pSS44 gene contents.

The IS*1414* presence highlighted the "deep sequencing" technology limits versus the Sanger method. Thus, 454 sequencing was required to be paired with Sanger technology. In fact, the high-throughput sequencing produces reads that are shorter than IS*1414* and assembly cannot align sequences ending in repeated regions. So far, 207 genome gaps remain to be closed (corresponding to less than 5 % of the total estimated length). It is highly possible that most, if not all, the inter-scaffold gap still "open" in the chromosome, and the two on the pSS44 plasmid, would correspond to further *IS1414* insertion loci.

Previously, data from Uzzau and coll. indicated a large content (over 30 copy number) of IS*1414* within serotype Abortusovis genome, few of which were also characterised up to the insertion site. Data collected in this research work, have demonstrated that IS*1414* copy number is much higher (approximatively 100) and that they differ for integrity and for the presence of the intragenic ORF *astA*.

Among the possible effects of such impressive amount of IS copy number, rarely observed in other bacteria, it is expected an enhancement of homologous intramolecular (chromosomal or plasmidic) recombination events and, furthermore, a number of structural and/or polar mutation scattered throughout the genome. Both predictable events have been demonstrated with the present research and, all together, point out toward a high level of genome plasticity and bacterial population adaptability.

Evidence of IS*1414*-mediated intramolecular recombination (physical rearrangements) were provided by direct sequencing (see below SPI-1 and SPI-3 comments on sequence analysis), but also by macrorestriction analysis with the rare cutter *Xba*I endonuclease. In especially, all strains lacking IS*1414* appeared to belong to a unique pulsotype, suggesting a monophyletic character and a lower recombinogenicity compared to more polymorphic pulsotypes occurring in IS*1414*$^{+}$ strains. Genome plasticity is a "mark" of other host-adapted and host-restricted pathogen, including serotype Typhi. Moreover, these classes of pathogens (including *Mycobacterium spp.*) often present a genetic degradation (number of loss of gene function), either by deletion or pseudogenes selection (Rondini *et al.*, 2007).

In addition, in this study, it has been demonstrated that one of such IS*1414* insertion is consistently located in an intergenic region (*iagB-sptP*) within SPI-1, where the integrated IS might trigger a polar mutation on upstream operon *hila-iagB*, a key player locus on bacterial intracellular uptake (*iagB*) and on master regulation of invasion effector expression (*hilA*) (Maithreye and Mande, 2007; Miras *et al.*, 1995).

IS*1414* encodes enteroaggregative *E. coli* heat-stable toxin 1 (EAST1), a small protein (38 amino acids) first discovered in EAEC *E. coli*. This protein shows significant homology with the enterotoxic domain of heat-stable enterotoxin a (STa) of enterotoxigenic (ETEC) *E. coli* and with guanylin, a mammalian analog of STa. Unlike STa, which requires six cysteines and three disulfide linkages for full biological activity, both EAST1 and guanylin contain four cysteine residues. EAST1 is predicted to stimulate the parti-

culate form of guanylate cyclase (GC) through the same receptor-binding region as STa and guanylin. While GC stimulation results in cGMP raise, EAST1 pathogenicity mechanisms have not been elucidated yet. So far, two variants are identified: EAST1 17-2 and 08-42 (encoded by different *E. coli* strains) with a single amino acid change (T to A) and different virulence levels. In fact, EAST1 08-42 is associated with diarrhea when administered to volunteers, contrarily to 17-2 variants. Sequencing data showed that SS44 EAST1 is the 17-2 variants, and this result is compatible with the non-diarrheagenic serotype Abortusovis (Veilleux and Dubreuil, 2006; Veilleux*, et al.*, 2008).

Alignment between EAST1 and guanylin or uroguanylin (GC-C agonists) showed similarity (identity of 5 or 8 aminoacids and similarity 19 or 16 amino acids, respectively). Therefore, EAST1 could bind GC-C, or another isoform with different physiologically effects. In fact, experimental evidences reported in this thesis induce to speculate on an effect associated to the inflammation.

A very large amount of sequence data was made available by 454 sequencing. Thus, it has been necessary to choose an initial group of loci, as the first to be analysed, in respect to possible homology to, or divergence from, other related sequences in host-adapted or ubiquitous *Salmonella* serotypes. *Salmonella* Pathogenicity Islands play a key role played in pathogenesis and they have been acquired during early evolution process of the genus *Salmonella*, that turns into a consistent presence within all subspecies of *S. enterica*. Then a major effort has been dedicated to SPI comparative analysis. Among serotype Abortusovis SPIs, SPI-1 and SPI-3 showed to be those with

the most interesting genetic rearrangement. SPI-1 has a pivotal role during intestinal mucosa invasion and inflammation. Strikingly, these biological and clinical traits of *Salmonella* infection are quite different when the animal host (i.e., ovine) is infected by SAO rather than by STM. Therefore, SPI-1 rearrangement in SAO is not surprising, but has never observed earlier in any other serotypes, including those with low inflammatory power (Figure 22). In addition to major physical reorganisation, SPI-1 gene *sptP* harbours a nonsense mutation that creates an early stop codon. Downstream, *sptP* coding RNA might origin to a polypeptide, corresponding to the last 438 amino acids of STM SptP. SptP is proposed to be required to antagonised SopE2 (or SopE) actin rearrangement activity. Such function, related to invasion and inflammation, might be deficient or stronger in the truncated SptP of serotype Abortusovis.

In serotype Typhimurium, SPI-3, previously identified as one of the major loci involved in *Salmonella* pathogenesis, carries several genes including *sugR* and *rhuM*, that lack in a number of other serotypes, suggesting a multistep gene acquisition during pathogen evolution. This work showed that, in addition to other serotypes (i.e., serotypes Bovismorbificans, Infantis, and Derby), Abortusovis did not acquired these genes, possibly because of advantageous for this serotype fitness and pathogenicity. Yet, sequencing of serotype Abortusovis SPI-3, revealed that as for SPI-1, insertion of IS*1414* might have mediated a major rearrangement by means of homologous recombination with an IS*1414* copy located elsewhere (see also model of SPI-1 rearrangement, Figure 22).

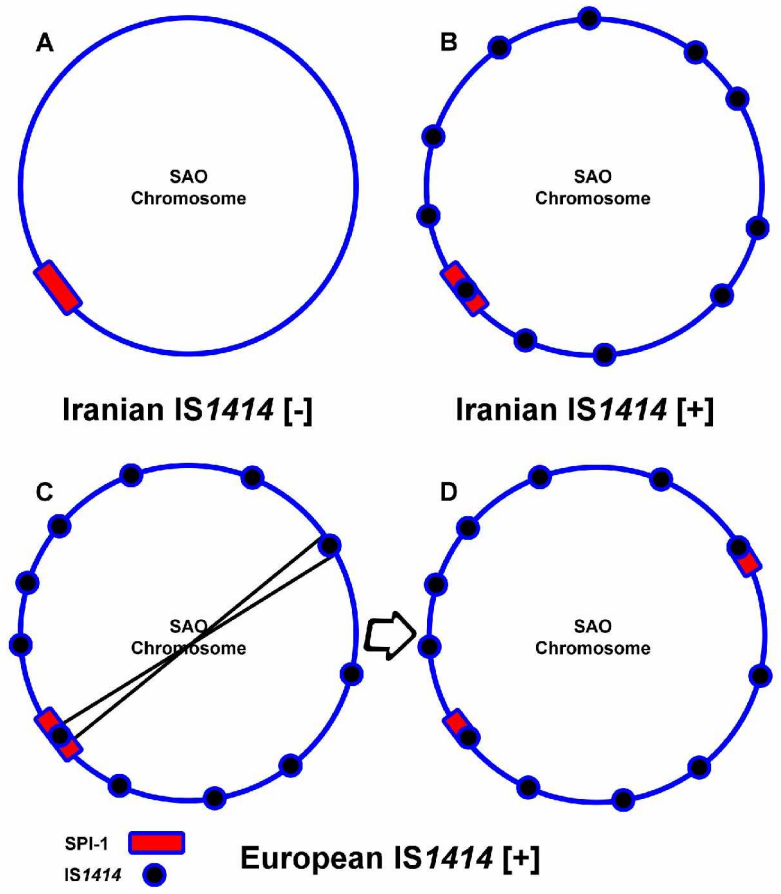

Figure 22. The scheme represents a hypothetical rearrangement in SPI-1 mediated by IS1414. A and B) An Iranian Abortusovis chromosome IS*1414* negative and positive, respectively. C) A European Abortusovis chromosome, before and after (D) a rearrangement. It is hypothesised a recombination between the insertion sequence inside the SPI-1 and another IS. The result is the current genomic structure. Also SPI-3 and other loci could be recombined with this pattern.

While waiting for a complete genome sequence and finished chromosomal structure, research activities were dedicated to an easier task: the pSS44 sequencing.

Firstly, the assembly was accomplished throughout two technologies, and our approach resulted correct. In fact, using a combination of Sanger 'traditional' sequenc-

ing and 454 'high-throughput' sequencing, the plasmid map has been almost completed

The plasmid harbours at least five IS*1414*, and analysis with SAO strains IS*1414*-negative showed large structural rearrangements (as for SPI-1 and SPI-3). Yet, pSS44 map revealed acquisition of a wide locus, corresponding to the scaffold #016. The scaffold, with a region homologue with other Salmonellae, presents IS*1414* insertions in the extremities. Also, scaffold GC-compositional strand bias (GC skew=(C-G)/(C+G)) is strongly different from the plasmidic backbone. GC bias inverts between replication origin and termination in bacterial chromosomes because numbers of G and T residues in the leading strand of these genomes exceed those of C and A (Fujimori *et al.*, 2005). Since, bacterial chromosomes display GC bias essentially constant, with the exception of origin and termination, scaffold #016 could have been "recently" transferred from an "external" chromosome.

Previous works (Uzzau *et al.*, personal communication) argued a plasmidic involvement in the inflammatory process in ovine. In fact, even though serotype Abortusovis is not enteritogen, SU40 strain (SS44 plasmid cured) is associated with inflammatory stimulus. To unravel this phenomenon, pSS44 digestion fragments were cloned into a low copy number vector (pJW4303, CN about 10). Each recombinant plasmid was used to transform SU40 cells. All strains (wild type included) induced swelling by *in vivo* experiment, with the exception of the strain carrying pDB108 plasmid, with a pSS44 fragment 7 kbp in length. The pDB108 insert was characterised and coding re-

gions analysed. pSS44 ORF analysis demonstrated that only one IS*1414* encoding an intact *astA* gene.It is noteworthy that the unique coding IS*1414* belong to this plasmid portion. The presence of supernumerary *astA* genes may affect the host response to EAST1. In fact, the vector employed in the experiment had a tenfold CN with respect to pSS44 plasmid virulence. Experimental evidences of a reciprocal association of inflammation and EAST1 gene copy number, suggest an EAST1 role in inflammation modulation.

Conversely, THP-1 infection experiments do not have demonstrated a direct link between plasmid encoded effector(s) (i.e., EAST1) and NF-κB activity, but this model does not exclude a possible role of EAST1.

In conclusion, this work drew attention to some structural features of a *Salmonella* host-restricted genome, achieved by high-throughput sequencing. One of the main points was the role of the IS*1414* insertion sequence, in particularly in DNA rearrangements, in acquisition of exogenous sequences. Furthermore it is interesting the possible effects, so far ravelled, of a toxin (the guanylin-like EAST1 polipeptide). In particular, EAST1 could induce a host response associated with the inflammation, emphasizing a crucial aspect of the non-enteritogen serotype Abortusovis.

This work brought a number of interesting future tasks and perspectives: i) complete sequence map, chromosomal and plasmidic; ii) achievement of all IS*1414* position throughout the genome; iii) characterisation of all IS*1414* insertion dependent

rearrangements and polar effect on adjacent genes; iv) obtain insights in the role of EAST1 during host infection.

Chapter 6

REFERENCES

Abe, A., Matsuzawa, T., Kuwae, A., (2005). Type-III effectors: sophisticated bacterial virulence factors. C R Biol. 328, 413-428.
Abrahams, G.L., Hensel, M., (2006). Manipulating cellular transport and immune responses: dynamic interactions between intracellular Salmonella enterica and its host cells. Cell Microbiol. 8, 728-737.
Aguirre, A., Cabeza, M.L., Spinelli, S.V., McClelland, M., Garcia Vescovi, E., Soncini, F.C., (2006). PhoP-induced genes within Salmonella pathogenicity island 1. J Bacteriol. 188, 6889-6898.
Altier, C., (2005). Genetic and environmental control of salmonella invasion. J Microbiol. 43 Spec No, 85-92.
Amavisit, P., Lightfoot, D., Browning, G.F., Markham, P.F., (2003). Variation between pathogenic serovars within Salmonella pathogenicity islands. J Bacteriol. 185, 3624-3635.
Artis, D., (2008). Epithelial-cell recognition of commensal bacteria and maintenance of immune homeostasis in the gut. Nat Rev Immunol. 8, 411-420.
Ausubel, F.M., (2001). Current protocols in molecular biology, J. Wiley, New York.
Bacciu, D., Falchi, G., Spazziani, A., Bossi, L., Marogna, G., Leori, G.S., Rubino, S.*, et al.*, (2004). Transposition of the heat-stable toxin astA gene into a gifsy-2-related prophage of Salmonella enterica serovar Abortusovis. J Bacteriol. 186, 4568-4574.
Baumler, A.J., Tsolis, R.M., Ficht, T.A., Adams, L.G., (1998). Evolution of host adaptation in Salmonella enterica. Infect Immun. 66, 4579-4587.
Biedzka-Sarek, M., El Skurnik, M., (2006). How to outwit the enemy: dendritic cells face Salmonella. APMIS. 114, 589-600.
Blanc-Potard, A.B., Solomon, F., Kayser, J., Groisman, E.A., (1999). The SPI-3 pathogenicity island of Salmonella enterica. J Bacteriol. 181, 998-1004.
Boddicker, J.D., Knosp, B.M., Jones, B.D., (2003). Transcription of the Salmonella invasion gene activator, hilA, requires HilD activation in the absence of negative regulators. J Bacteriol. 185, 525-533.
Boyd, E.F., Hartl, D.L., (1998). Salmonella virulence plasmid. Modular acquisition of the spv virulence region by an F-plasmid in Salmonella enterica subspecies I and insertion into the chromosome of subspecies II, IIIa, IV and VII isolates. Genetics. 149, 1183-1190.
Brown, N.F., Vallance, B.A., Coombes, B.K., Valdez, Y., Coburn, B.A., Finlay, B.B., (2005). Salmonella pathogenicity island 2 is expressed prior to penetrating the intestine. PLoS Pathog. 1, e32.
Camacho, E.M., Casadesus, J., (2005). Regulation of traJ transcription in the Salmonella virulence plasmid by strand-specific DNA adenine hemimethylation. Mol Microbiol. 57, 1700-1718.
Camacho, E.M., Serna, A., Casadesus, J., (2005a). Regulation of conjugal transfer by Lrp and Dam methylation in plasmid R100. Int Microbiol. 8, 279-285.
Camacho, E.M., Serna, A., Madrid, C., Marques, S., Fernandez, R., de la Cruz, F., Juarez, A.*, et al.*, (2005b). Regulation of finP transcription by DNA adenine methylation in the virulence plasmid of Salmonella enterica. J Bacteriol. 187, 5691-5699.
Chiu, C.H., Tang, P., Chu, C., Hu, S., Bao, Q., Yu, J., Chou, Y.Y.*, et al.*, (2005). The genome sequence of Salmonella enterica serovar Choleraesuis, a highly invasive and resistant zoonotic pathogen. Nucleic Acids Res. 33, 1690-1698.
Chu, C., Chiu, C.H., (2006). Evolution of the virulence plasmids of non-typhoid Salmonella and its association with antimicrobial resistance. Microbes Infect. 8, 1931-1936.

Chu, C., Chiu, C.H., Chu, C.H., Ou, J.T., (2002). Nucleotide and amino acid sequences of oriT-traM-traJ-traY-traA-traL regions and mobilization of virulence plasmids of Salmonella enterica serovars enteritidis, gallinarum-pullorum, and typhimurium. J Bacteriol. 184, 2857-2862.
Chu, C., Feng, Y., Chien, A.C., Hu, S., Chu, C.H., Chiu, C.H., (2008). Evolution of genes on the Salmonella Virulence plasmid phylogeny revealed from sequencing of the virulence plasmids of S. enterica serotype Dublin and comparative analysis. Genomics.
Cornelis, G.R., (2006). The type III secretion injectisome. Nat Rev Microbiol. 4, 811-825.
Crosa, J.H., Brenner, D.J., Ewing, W.H., Falkow, S., (1973). Molecular relationships among the Salmonelleae. J Bacteriol. 115, 307-315.
Deng, W., Liou, S.R., Plunkett, G., 3rd, Mayhew, G.F., Rose, D.J., Burland, V., Kodoyianni, V., *et al.*, (2003). Comparative genomics of Salmonella enterica serovar Typhi strains Ty2 and CT18. J Bacteriol. 185, 2330-2337.
Dieye, Y., Dyszel, J.L., Kader, R., Ahmer, B.M., (2007). Systematic analysis of the regulation of type three secreted effectors in Salmonella enterica serovar Typhimurium. BMC Microbiol. 7, 3.
Dionisi, A.M., Carattoli, A., Luzzi, I., Magistrali, C., Pezzotti, G., (2006). Molecular genotyping of Salmonella enterica Abortusovis by pulsed field gel electrophoresis. Vet Microbiol. 116, 217-223.
Drecktrah, D., Knodler, L.A., Howe, D., Steele-Mortimer, O., (2007). Salmonella trafficking is defined by continuous dynamic interactions with the endolysosomal system. Traffic. 8, 212-225.
Ehrbar, K., Friebel, A., Miller, S.I., Hardt, W.D., (2003). Role of the Salmonella pathogenicity island 1 (SPI-1) protein InvB in type III secretion of SopE and SopE2, two Salmonella effector proteins encoded outside of SPI-1. J Bacteriol. 185, 6950-6967.
Ellermeier, J.R., Slauch, J.M., (2007). Adaptation to the host environment: regulation of the SPI1 type III secretion system in Salmonella enterica serovar Typhimurium. Curr Opin Microbiol. 10, 24-29.
Figueroa-Bossi, N., Bossi, L., (1999). Inducible prophages contribute to Salmonella virulence in mice. Mol Microbiol. 33, 167-176.
Fink, S.L., Cookson, B.T., (2005). Apoptosis, pyroptosis, and necrosis: mechanistic description of dead and dying eukaryotic cells. Infect Immun. 73, 1907-1916.
Fink, S.L., Cookson, B.T., (2007). Pyroptosis and host cell death responses during Salmonella infection. Cell Microbiol. 9, 2562-2570.
Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., *et al.*, (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science. 269, 496-512.
Folkesson, A., Advani, A., Sukupolvi, S., Pfeifer, J.D., Normark, S., Lofdahl, S., (1999). Multiple insertions of fimbrial operons correlate with the evolution of Salmonella serovars responsible for human disease. Mol Microbiol. 33, 612-622.
Fujimori, S., Washio, T., Tomita, M., (2005). GC-compositional strand bias around transcription start sites in plants and fungi. BMC Genomics. 6, 26.
Galan, J.E., Wolf-Watz, H., (2006). Protein delivery into eukaryotic cells by type III secretion machines. Nature. 444, 567-573.
Goldberg, S.M., Johnson, J., Busam, D., Feldblyum, T., Ferriera, S., Friedman, R., Halpern, A., *et al.*, (2006). A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. Proc Natl Acad Sci U S A. 103, 11240-11245.
Gorvel, J.P., Meresse, S., (2001). Maturation steps of the Salmonella-containing vacuole. Microbes Infect. 3, 1299-1303.

Grassl, G.A., Finlay, B.B., (2008). Pathogenesis of enteric Salmonella infections. Curr Opin Gastroenterol. 24, 22-26.
Grimont, P.A.D., Weill, F.X., (2007). Antigenic formulae of the Salmonella serovars, WHO Collaborating Centre for Reference and Research on Salmonella.
Groisman, E.A., (2001). The pleiotropic two-component regulatory system PhoP-PhoQ. J Bacteriol. 183, 1835-1842.
Guiney, D.G., Lesnick, M., (2005). Targeting of the actin cytoskeleton during infection by Salmonella strains. Clin Immunol. 114, 248-255.
Guo, A., Lasaro, M.A., Sirard, J.C., Kraehenbuhl, J.P., Schifferli, D.M., (2007). Adhesin-dependent binding and uptake of Salmonella enterica serovar Typhimurium by dendritic cells. Microbiology. 153, 1059-1069.
Hall, N., (2007). Advanced sequencing technologies and their wider impact in microbiology. J Exp Biol. 210, 1518-1525.
Haraga, A., Ohlson, M.B., Miller, S.I., (2008). Salmonellae interplay with host cells. Nat Rev Microbiol. 6, 53-66.
Harris, T.D., Buzby, P.R., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., Causey, M., *et al.*, (2008). Single-molecule DNA sequencing of a viral genome. Science. 320, 106-109.
Holt, K.E., Parkhill, J., Mazzoni, C.J., Roumagnac, P., Weill, F.X., Goodhead, I., Rance, R., *et al.*, (2008). High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi. Nat Genet. 40, 987-993.
Hughes, D., (2000). Evaluating genome dynamics: the constraints on rearrangements within bacterial genomes. Genome Biol. 1, REVIEWS0006.
Iliev, I.D., Matteoli, G., Rescigno, M., (2007). The yin and yang of intestinal epithelial cells in controlling dendritic cell function. J Exp Med. 204, 2253-2257.
Jones, B.D., (2005). Salmonella invasion gene regulation: a story of environmental awareness. J Microbiol. 43 Spec No, 110-117.
Kiss, T., Morgan, E., Nagy, G., (2007). Contribution of SPI-4 genes to the virulence of Salmonella enterica. FEMS Microbiol Lett. 275, 153-159.
Kothapalli, S., Nair, S., Alokam, S., Pang, T., Khakhria, R., Woodward, D., Johnson, W., *et al.*, (2005). Diversity of genome structure in Salmonella enterica serovar Typhi populations. J Bacteriol. 187, 2638-2650.
Kubori, T., Galan, J.E., (2003). Temporal regulation of salmonella virulence effector function by proteasome-dependent protein degradation. Cell. 115, 333-342.
Lee, F.K., Morris, C., Hackett, J., (2006). The Salmonella enterica serovar Typhi Vi capsule and self-association pili share controls on expression. FEMS Microbiol Lett. 261, 41-46.
Liao, A.P., Petrof, E.O., Kuppireddi, S., Zhao, Y., Xia, Y., Claud, E.C., Sun, J., (2008). Salmonella type III effector AvrA stabilizes cell tight junctions to inhibit inflammation in intestinal epithelial cells. PLoS ONE. 3, e2369.
Lilic, M., Galkin, V.E., Orlova, A., VanLoock, M.S., Egelman, E.H., Stebbins, C.E., (2003). Salmonella SipA polymerizes actin by stapling filaments with nonglobular protein arms. Science. 301, 1918-1921.
Lin, S.L., Le, T.X., Cowen, D.S., (2003). SptP, a Salmonella typhimurium type III-secreted protein, inhibits the mitogen-activated protein kinase pathway by inhibiting Raf activation. Cell Microbiol. 5, 267-275.
Liu, G.-R., Liu, W.-Q., Johnston, R.N., Sanderson, K.E., Li, S.-X., Liu, S.-L., (2006). Genome Plasticity and ori-ter Rebalancing in Salmonella typhi. Mol Biol Evol. 23, 365-371.
Liu, S.L., Hessel, A., Sanderson, K.E., (1993). Genomic mapping with I-Ceu I, an intron-encoded endonuclease specific for genes for ribosomal RNA, in Salmonella spp., Escherichia coli, and other bacteria. Proc Natl Acad Sci U S A. 90, 6874-6878.

Liu, W.Q., Liu, G.R., Li, J.Q., Xu, G.M., Qi, D., He, X.Y., Deng, J., *et al.*, (2007). Diverse genome structures of Salmonella paratyphi C. BMC Genomics. 8, 290.
Lober, S., Jackel, D., Kaiser, N., Hensel, M., (2006). Regulation of Salmonella pathogenicity island 2 genes by independent environmental signals. Int J Med Microbiol. 296, 435-447.
Ly, K.T., Casanova, J.E., (2007). Mechanisms of Salmonella entry into host cells. Cell Microbiol. 9, 2103-2111.
Maithreye, R., Mande, S.S., (2007). Modelling of the regulation of the hilA promoter of type three secretion system of Salmonella enterica serovar Typhimurium. Syst Synth Biol. 1, 129-137.
Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., *et al.*, (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature. 437, 376-380.
McClelland, M., Sanderson, K.E., Clifton, S.W., Latreille, P., Porwollik, S., Sabo, A., Meyer, R., *et al.*, (2004). Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of Salmonella enterica that cause typhoid. Nat Genet. 36, 1268-1274.
McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Porwollik, S., *et al.*, (2001). Complete genome sequence of Salmonella enterica serovar Typhimurium LT2. Nature. 413, 852-856.
McVeigh, A., Fasano, A., Scott, D.A., Jelacic, S., Moseley, S.L., Robertson, D.C., Savarino, S.J., (2000). IS1414, an Escherichia coli insertion sequence with a heat-stable enterotoxin gene embedded in a transposase-like gene. Infect Immun. 68, 5710-5715.
Medini, D., Serruto, D., Parkhill, J., Relman, D.A., Donati, C., Moxon, R., Falkow, S., *et al.*, (2008). Microbiology in the post-genomic era. Nat Rev Microbiol. 6, 419-430.
Miras, I., Hermant, D., Arricau, N., Popoff, M.Y., (1995). Nucleotide sequence of iagA and iagB genes involved in invasion of HeLa cells by Salmonella enterica subsp. enterica ser. Typhi. Res Microbiol. 146, 17-20.
Nikbakht, G.H., Raffatellu, M., Uzzau, S., Tadjbakhsh, H., Rubino, S., (2002). IS200 fingerprinting of Salmonella enterica serotype Abortusovis strains isolated in Iran. Epidemiol Infect. 128, 333-336.
Paiva de Sousa, C., Dubreuil, J.D., (2001). Distribution and expression of the astA gene (EAST1 toxin) in Escherichia coli and Salmonella. Int J Med Microbiol. 291, 15-20.
Park, J., Schulz, S., Waldman, S.A., (2000). Intestine-specific activity of the human guanylyl cyclase C promoter is regulated by Cdx2. Gastroenterology. 119, 89-96.
Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Wain, J., Churcher, C., *et al.*, (2001). Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18. Nature. 413, 848-852.
Perez, J.C., Groisman, E.A., (2007). Acid pH activation of the PmrA/PmrB two-component regulatory system of Salmonella enterica. Mol Microbiol. 63, 283-293.
Peters, T.M., Maguire, C., Threlfall, E.J., Fisher, I.S., Gill, N., Gatto, A.J., (2003). The Salm-gene project - a European collaboration for DNA fingerprinting for. Euro Surveill. 8, 46-50.
Pickard, D., Wain, J., Baker, S., Line, A., Chohan, S., Fookes, M., Barron, A., *et al.*, (2003). Composition, acquisition, and distribution of the Vi exopolysaccharide-encoding Salmonella enterica pathogenicity island SPI-7. J Bacteriol. 185, 5055-5065.
Prost, L.R., Miller, S.I., (2008). The Salmonellae PhoQ sensor: mechanisms of detection of phagosome signals. Cell Microbiol. 10, 576-582.
Rabsch, W., Andrews, H.L., Kingsley, R.A., Prager, R., Tschape, H., Adams, L.G., Baumler, A.J., (2002). Salmonella enterica serotype Typhimurium and its host-adapted variants. Infect Immun. 70, 2249-2255.

Ran He., (2005). Rapid and efficient generation of pcr templates from *Escherichia coli*, *Saccharomyces* cerevisiae and *oryza sativa* using a microwave and by boiling. Journal of Rapid Methods & Automation in Microbiology. 13, 19-28.
Rescigno, M., (2002). Dendritic cells and the complexity of microbial infection. Trends Microbiol. 10, 425-461.
Rescigno, M., Chieppa, M., (2005). Gut-level decisions in peace and war. Nat Med. 11, 254-255.
Rescigno, M., Urbano, M., Valzasina, B., Francolini, M., Rotta, G., Bonasio, R., Granucci, F.*, et al.*, (2001). Dendritic cells express tight junction proteins and penetrate gut epithelial monolayers to sample bacteria. Nat Immunol. 2, 361-367.
Rondini, S., Kaser, M., Stinear, T., Tessier, M., Mangold, C., Dernick, G., Naegeli, M.*, et al.*, (2007). Ongoing genome reduction in Mycobacterium ulcerans. Emerg Infect Dis. 13, 1008-1015.
Rychlik, I., Gregorova, D., Hradecka, H., (2006). Distribution and function of plasmids in Salmonella enterica. Vet Microbiol. 112, 1-10.
Sanger, F., Nicklen, S., Coulson, A.R., (1977). DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 74, 5463-5467.
Santos, R.L., Tsolis, R.M., Baumler, A.J., Adams, L.G., (2003). Pathogenesis of Salmonella-induced enteritis. Braz J Med Biol Res. 36, 3-12.
Savarino, S.J., Fasano, A., Robertson, D.C., Levine, M.M., (1991). Enteroaggregative Escherichia coli elaborate a heat-stable enterotoxin demonstrable in an in vitro rabbit intestinal model. J Clin Invest. 87, 1450-1455.
Schiaffino, A., Beuzon, C.R., Uzzau, S., Leori, G., Cappuccinelli, P., Casadesus, J., Rubino, S., (1996). Strain typing with IS200 fingerprints in Salmonella abortusovis. Appl Environ Microbiol. 62, 2375-2380.
Schlumberger, M.C., Hardt, W.D., (2006). Salmonella type III secretion effectors: pulling the host cell's strings. Curr Opin Microbiol. 9, 46-54.
Service, R.F., (2006). Gene sequencing. The race for the $1000 genome. Science. 311, 1544-1546.
Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D.*, et al.*, (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. Science. 309, 1728-1732.
Steele-Mortimer, O., (2008). The Salmonella-containing vacuole: moving with the times. Curr Opin Microbiol. 11, 38-45.
Stiens, M., Becker, A., Bekel, T., Godde, V., Goesmann, A., Niehaus, K., Schneiker-Bekel, S.*, et al.*, (2008). Comparative genomic hybridisation and ultrafast pyrosequencing revealed remarkable differences between the Sinorhizobium meliloti genomes of the model strain Rm1021 and the field isolate SM11. J Biotechnol. 136, 31-37.
Suarez, M., Russmann, H., (1998). Molecular mechanisms of Salmonella invasion: the type III secretion system of the pathogenicity island 1. Int Microbiol. 1, 197-204.
Thompson, A., Rolfe, M.D., Lucchini, S., Schwerk, P., Hinton, J.C., Tedin, K., (2006). The bacterial signal molecule, ppGpp, mediates the environmental regulation of both the invasion and intracellular virulence gene programs of Salmonella. J Biol Chem. 281, 30112-30121.
Thomson, N.R., Clayton, D.J., Windhorst, D., Vernikos, G., Davidson, S., Churcher, C., Quail, M.A.*, et al.*, (2008). Comparative genome analysis of Salmonella Enteritidis PT4 and Salmonella Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. Genome Res.

Tukel, C., Raffatellu, M., Chessa, D., Wilson, R.P., Akcelik, M., Baumler, A.J., (2006). Neutrophil influx during non-typhoidal salmonellosis: who is in the driver's seat? FEMS Immunol Med Microbiol. 46, 320-329.
Uzzau, S., Brown, D.J., Wallis, T., Rubino, S., Leori, G., Bernard, S., Casadesus, J., *et al.*, (2000a). Host adapted serotypes of Salmonella enterica. Epidemiol Infect. 125, 229-255.
Uzzau, S., Gulig, P.A., Paglietti, B., Leori, G., Stocker, B.A., Rubino, S., (2000b). Role of the Salmonella abortusovis virulence plasmid in the infection of BALB/c mice. FEMS Microbiol Lett. 188, 15-18.
Uzzau, S., Marogna, G., Leori, G.S., Curtiss, R., 3rd, Schianchi, G., Stocker, B.A., Rubino, S., (2005). Virulence attenuation and live vaccine potential of aroA, crp cdt cya, and plasmid-cured mutants of Salmonella enterica serovar Abortusovis in mice and sheep. Infect Immun. 73, 4302-4308.
Valdezate, S., Astorga, R., Herrera-Leon, S., Perea, A., Usera, M.A., Huerta, B., Echeita, A., (2007). Epidemiological tracing of Salmonella enterica serotype Abortusovis from Spanish ovine flocks by PFGE fingerprinting. Epidemiol Infect. 135, 695-702.
Veilleux, S., Dubreuil, J.D., (2006). Presence of Escherichia coli carrying the EAST1 toxin gene in farm animals. Vet Res. 37, 3-13.
Veilleux, S., Holt, N., Schultz, B.D., Dubreuil, J.D., (2008). Escherichia coli EAST1 toxin toxicity of variants 17-2 and O 42. Comp Immunol Microbiol Infect Dis. 31, 567-578.
Waterman, S.R., Holden, D.W., (2003). Functions and effectors of the Salmonella pathogenicity island 2 type III secretion system. Cell Microbiol. 5, 501-511.
Wick, M.J., (2007). Monocyte and dendritic cell recruitment and activation during oral Salmonella infection. Immunol Lett. 112, 68-74.
Wicker, T., Schlagenhauf, E., Graner, A., Close, T.J., Keller, B., Stein, N., (2006). 454 sequencing put to the test using the complex genome of barley. BMC Genomics. 7, 275.
Wilson, J.W., Nickerson, C.A., (2006). A new experimental approach for studying bacterial genomic island evolution identifies island genes with bacterial host-specific expression patterns. BMC Evol Biol. 6, 2.
Winnen, B., Schlumberger, M.C., Sturm, A., Schupbach, K., Siebenmann, S., Jenny, P., Hardt, W.D., (2008). Hierarchical effector protein transport by the Salmonella Typhimurium SPI-1 type III secretion system. PLoS ONE. 3, e2178.
Wood, M.W., Jones, M.A., Watson, P.R., Hedges, S., Wallis, T.S., Galyov, E.E., (1998). Identification of a pathogenicity island required for Salmonella enteropathogenicity. Mol Microbiol. 29, 883-891.
Ye, Z., Petrof, E.O., Boone, D., Claud, E.C., Sun, J., (2007). Salmonella effector AvrA regulation of colonic epithelial cell inflammation by deubiquitination. Am J Pathol. 171, 882-892.
Zhang, S., Kingsley, R.A., Santos, R.L., Andrews-Polymenis, H., Raffatellu, M., Figueiredo, J., Nunes, J., *et al.*, (2003). Molecular pathogenesis of Salmonella enterica serotype typhimurium-induced diarrhea. Infect Immun. 71, 1-12.
Zhang, S., Santos, R.L., Tsolis, R.M., Stender, S., Hardt, W.D., Baumler, A.J., Adams, L.G., (2002). The Salmonella enterica serotype typhimurium effector proteins SipA, SopA, SopB, SopD, and SopE2 act in concert to induce diarrhea in calves. Infect Immun. 70, 3843-3855.

Acknowledgements

Firstly, I would like to thank my supervisor Prof. Sergio Uzzau for his guidance during these three years and for trying to teach me how to do research. Special thanks to Donatella Bacciu for her patience and valuable experience. My gratitude to my colleagues Francesca Campesi (for OMP-ELISA), Manuela Murgia and Bianca Paglietti for their precious advices about the PFGE. Also, many thanks to Dr. de Bellis and his colleagues Dr. Giorgio Corti, Dr. Michele Iacono and Dr. Roberta Bordoni for giving me the great opportunity to learn about bioinformatics, and for their huge work with the sequences.

Finally, I am pleased to thank my colleague and dear friend Elisa, who supports me with her English skills and encouragement throughout my PhD studies.