

Research article

Open Access

Comparative genomic analysis uncovers 3 novel loci encoding type six secretion systems differentially distributed in *Salmonella* serotypes

Carlos J Blondel¹, Juan C Jiménez¹, Inés Contreras¹ and Carlos A Santiviago^{*2}

Address: ¹Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, P.O. Box 174, Correo 22, Santiago, Chile and ²Programa de Microbiología y Micología, Instituto de Ciencias Biomédicas, Facultad de Medicina, Universidad de Chile, Av. Independencia 1027, Santiago, Chile

Email: Carlos J Blondel - cblondel@ciq.uchile.cl; Juan C Jiménez - jjromaguera@yahoo.es; Inés Contreras - icontrer@uchile.cl; Carlos A Santiviago* - csantiviago@gmail.com

* Corresponding author

Published: 4 August 2009

Received: 2 June 2009

BMC Genomics 2009, 10:354 doi:10.1186/1471-2164-10-354

Accepted: 4 August 2009

This article is available from: <http://www.biomedcentral.com/1471-2164/10/354>

© 2009 Blondel et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The recently described Type VI Secretion System (T6SS) represents a new paradigm of protein secretion in bacteria. A number of bioinformatic studies have been conducted to identify T6SS gene clusters in the available bacterial genome sequences. According to these studies, *Salmonella* harbors a unique T6SS encoded in the *Salmonella* Pathogenicity Island 6 (SPI-6). Since these studies only considered few *Salmonella* genomes, the present work aimed to identify novel T6SS loci by *in silico* analysis of every genome sequence of *Salmonella* available.

Results: The analysis of sequencing data from 44 completed or in progress *Salmonella* genome projects allowed the identification of 3 novel T6SS loci. These clusters are located in differentially-distributed genomic islands we designated SPI-19, SPI-20 and SPI-21, respectively. SPI-19 was identified in a subset of *S. enterica* serotypes including Dublin, Weltevreden, Agona, Gallinarum and Enteritidis. In the later, an internal deletion eliminated most of the island. On the other hand, SPI-20 and SPI-21 were restricted to *S. enterica* subspecies *arizonae* (IIIa) serotype 62:z4,z23:-. Remarkably, SPI-21 encodes a VgrG protein containing a C-terminal extension similar to S-type pyocins of *Pseudomonas aeruginosa*. This is not only the first evolved VgrG described in *Salmonella*, but also the first evolved VgrG including a pyocin domain described so far in the literature. In addition, the data indicate that SPI-6 T6SS is widely distributed in *S. enterica* and absent in serotypes Enteritidis, Gallinarum, Agona, Javiana, Paratyphi B, Virchow, IIIa 62:z4,z23:- and IIIb 61:l,v:1,5,(7). Interestingly, while some serotypes harbor multiple T6SS (Dublin, Weltevreden and IIIa 62:z4,z23:-) others do not encode for any (Enteritidis, Paratyphi B, Javiana, Virchow and IIIb 61:l,v:1,5,(7)). Comparative and phylogenetic analyses indicate that the 4 T6SS loci in *Salmonella* have a distinct evolutionary history. Finally, we identified an orphan Hcp-like protein containing the Hcp/COG3157 domain linked to a C-terminal extension. We propose to designate this and related proteins as "evolved Hcps".

Conclusion: Altogether, our data suggest that (i) the *Salmonella* T6SS loci were acquired by independent lateral transfer events and (ii) evolved to contribute in the adaptation of the serotypes to different lifestyles and environments, including animal hosts. Notably, the presence of an evolved VgrG protein related to pyocins suggests a novel role for T6SS in bacterial killing. Future studies on the roles of the identified T6SS loci will expand our knowledge on *Salmonella* pathogenesis and host specificity.

Background

The recently described Type VI Secretion System (T6SS) represents a new paradigm of protein secretion in bacteria maintaining pathogenic or symbiotic interactions with eukaryotic organisms (reviewed in references [1-4]). In this context, T6SSs have been linked to a variety of functions such as adherence, cytotoxicity, host-cell invasion, biofilm formation, survival within macrophages, and persistence within the host [2,4].

T6SS are encoded in loci initially known as IAHP (IcmF associated homologous protein) clusters. These gene clusters were characterized by the presence of approximately 15 ORFs surrounding a homolog of IcmF, a protein associated with type IV secretion in *Legionella pneumophila* [5-7]. In addition to the variable genetic architecture presented by IHAPs in several microbial genomes, low level of sequence identity between components of the system in different bacteria has hampered the identification of new T6SS loci. In spite of this, *in silico* analyses identified a set of 13 conserved proteins defined as the T6SS "core components" [8]. These components carry a distinct COG ID (Cluster of Orthologous Groups of proteins), most of which are unique for T6SS function [1,8]. In addition, there are 15 conserved accessory proteins widely distributed among T6SS loci that include specific transcriptional and post-transcriptional regulators [8].

Most T6SS core components correspond to structural elements of the secretion machinery. The DotU and IcmF homologs (COG3523 and COG3455, respectively) are conserved inner membrane proteins essential for secretion [9,10]. A role for IcmF and DotU homologs in membrane stabilization of the T6SS apparatus has been proposed based on observations made for homolog proteins involved in stabilization of the T4SS apparatus in *L. pneumophila* [6]. The ClpV homologs (COG0542) belong to the AAA+ family of ATPases and are hypothesized to energize the system, enabling protein secretion [9,11,12]. Recently, it has been reported that *V. cholerae* ClpV interacts through its N-terminal domain with a tubular structure formed by VipA (IglA, COG3516) and VipB (IglB, COG 3517), two conserved and essential T6SS components. Remodeling of VipA/VipB tubules by ClpV-mediated threading is crucial for type VI protein secretion [13]. In addition, it has been recently proposed that the VipA/VipB structure corresponds to a structural and possibly functional homolog of the tail sheath in bacteriophage T4, as both structures present similar dimensions, symmetry and overall organization [14].

Several other T6SS core components resemble proteins of bacteriophage origin. This is the case of the hemolysin coregulated protein Hcp (COG3157), the valine-glycine repeat protein VgrG (COG3501) and the gp25-like protein (COG3518), which are homologs to components of

the tail and baseplate in bacteriophage T4 [14,15]. Interestingly, both Hcp and VgrG appear to be structural and secreted components. Thus, the detection of both proteins in culture supernatants has become an indicator of T6SS functionality [10,16,17]. Some VgrG proteins known as "evolved VgrGs" present a C-terminal extension including "effector domains" that have been linked to a variety of functions, such as crosslinking of host actin, degradation of the peptidoglycan layer, and ADP-ribosylation of host proteins [4,16]. The evolved VgrGs have been identified in a limited number of bacteria and are usually encoded outside T6SS loci, scattered throughout the genome.

Following the initial characterization of IAHP clusters [5], several *in silico* analyses have identified T6SS loci in bacterial genome sequences available. In 2008, a screen for homologs to the conserved T6SS components VipA/IglA and VipB/IglB identified 37 T6SS loci in 29 bacteria [1]. Later that year, a similarity search for orthologs of T6SS components described in *Vibrio cholerae*, *Pseudomonas aeruginosa* and *Burkholderia mallei* identified T6SS loci in 42 bacterial species [18]. Even though both analyses greatly expanded the number of T6SS loci reported, they were based on the detection of homologs to T6SS components present in only three organisms, overlooking the existence of phylogenetically distant T6SS loci. A more recent study addressed this limitation by analyzing the presence of conserved protein domains rather than orthologous proteins, to identify 176 T6SS loci from 92 different bacteria [8].

The first report for a T6SS in *Salmonella* corresponds to the genetic characterization of the *Salmonella* Pathogenicity Island 6 (SPI-6), formerly known as SCI (*Salmonella enterica* centisome 7 island), which is adjacent to the tRNA-encoding gene *aspV* [19]. SPI-6 is a ~47 kb genomic island presenting a mosaic structure characterized by the presence of the *saf* and *tcf* fimbrial operons. Analysis of the region upstream of the fimbrial operons revealed numerous ORFs encoding putative periplasmic, outer membrane and secreted proteins, suggesting that SPI-6 encoded a novel secretion system [19]. After the classification of SPI-6 as an IAHP locus [5] and the discovery of T6SSs in *V. cholerae* and *P. aeruginosa* [9,20] it was clear that SPI-6 encoded a T6SS, although there is no experimental evidence on the functionality of this system. Conflicting reports associate SPI-6 T6SS with *Salmonella* pathogenesis. A *S. Typhimurium* mutant in STM0285 (IcmF homolog) presented an increased intracellular growth in macrophages and was hypervirulent in BALB/c mice [21]. On the other hand, mutations in STM0272 (ClpV homolog) and STM0291 (Rhs element) have been reported to cause a ~30% reduction in the ability of *S. Typhimurium* to replicate in macrophages [22]. In addition, overexpression of a dominant negative version of

STM0272 impaired the ability of *S. Typhimurium* to invade epithelial cells [12].

Despite the recent advances in our understanding of the structural and genomic organization of T6SS gene clusters, every *in silico* analysis performed so far to identify these loci in bacteria has taken into account a limited number of the currently available *Salmonella* genome sequences. Considering that many bacterial species harbor multiple T6SS in their genomes, it is plausible to think that there could be more than one T6SS in the genus *Salmonella* in addition to SPI-6 T6SS waiting to be uncovered. In the present work we performed a genome-wide *in silico* analysis of all currently available *Salmonella* genome sequences to identify T6SS loci. Our analysis revealed the presence of 3 novel T6SS gene clusters encoded in differentially-distributed genomic islands presenting distinctive evolutionary histories.

Results and Discussion

Identification of T6SS gene clusters in *Salmonella*

Representatives of the 13 T6SS core components recently defined by Boyer and coworkers [8] were used as baits to identify T6SS loci by sequential BLASTN, BLASTP and TBLASTX searches using all publicly available sequences from 44 *Salmonella* genome sequencing projects (completed or in progress). The data analyzed included both chromosome and plasmid sequences covering 24 different serotypes of *Salmonella enterica* (22 in subspecies I, 1 in subspecies IIIa and 1 in subspecies IIIb) (Additional file 1).

To maximize the power of the screen, representatives of each core component from 3 different organisms belonging to each of the different branches in phylogenetic trees previously defined for T6SS loci [1,8] were used as baits. The analysis revealed the presence of 3 novel T6SS loci in addition to the gene cluster encoded within SPI-6 (Table 1). The novel T6SS gene clusters are located in 3 genomic islands we have designated SPI-19, SPI-20 and SPI-21, respectively. These islands are differentially distributed among the *Salmonella* serotypes analyzed and contain each of the 13 T6SS core components described to date (Figure 1). Most *Salmonella* present a unique T6SS encoded either by SPI-6 or SPI-19. On the other hand, a limited number of *Salmonella* carry two T6SS, encoded either by SPI-6 and SPI-19 or by SPI-20 and SPI-21. Finally, no T6SS locus was identified in the available genomes of *S. enterica* subspecies *enterica* (I) serotypes Paratyphi B, Virchow and Javiana, and *S. enterica* subspecies *diarizonae* (IIIb) serotype 61:1,v:1,5,(7).

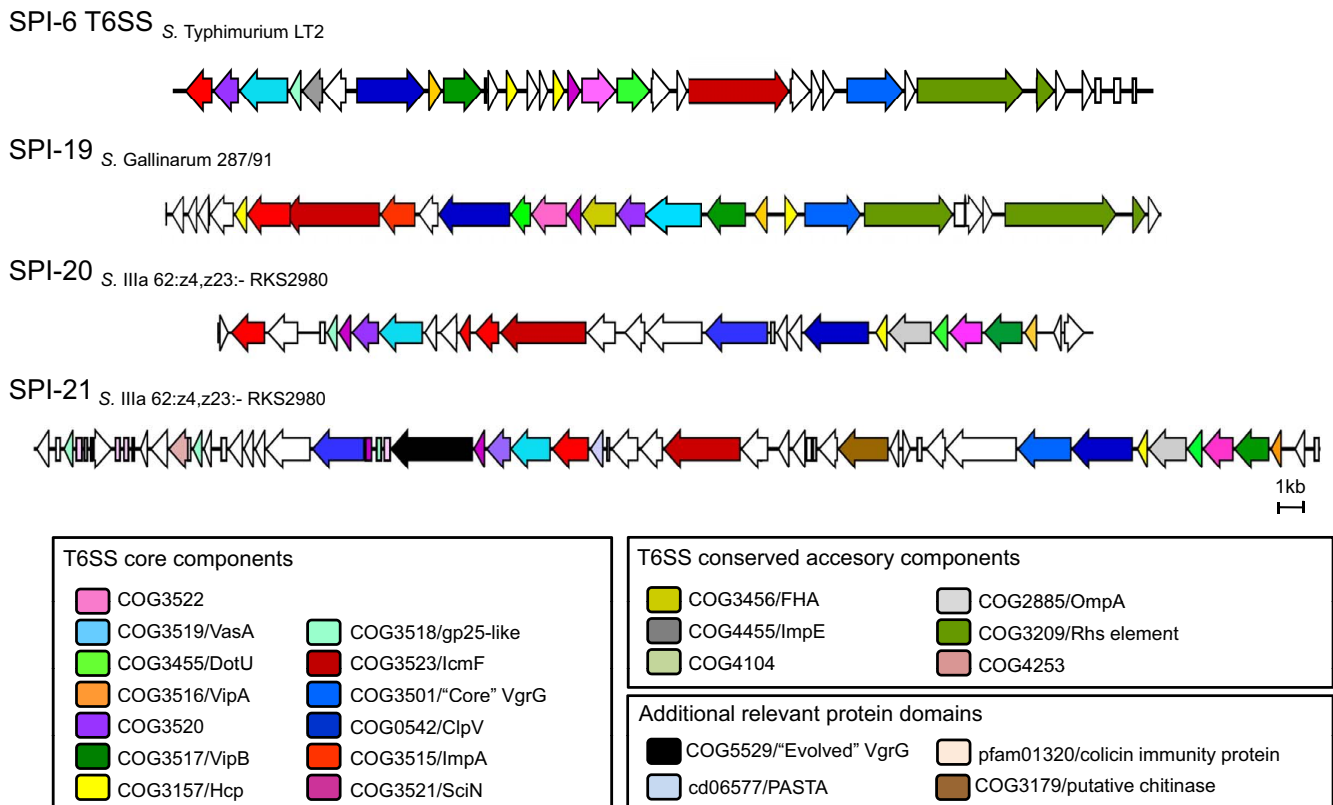
SPI-19: a second T6SS gene cluster in *S. enterica* subspecies *enterica* (I)

SPI-19 is a novel T6SS locus present in a subset of serotypes belonging to *S. enterica* subspecies *enterica* (I),

including Dublin, Weltevreden, Agona, Gallinarum and Enteritidis. Remarkably, serotypes Dublin and Weltevreden also carry the T6SS encoded in SPI-6, while serotype Enteritidis only presents a truncated version of SPI-19 (see below). The ORFs, product sizes, conserved protein domains and genome coordinates of SPI-19 genes in serotypes Dublin, Weltevreden, Agona, Gallinarum and Enteritidis are detailed in Additional file 2.

The island corresponds to a ~45 kb element located in the vicinity of a gene cluster encoding putative proteins involved in sugar transport and utilization (STM1127 to STM1133 in serotype Typhimurium), and is not directly linked to a tRNA-encoding gene (see below). Although the average G+C content of the island is similar to that of the whole genome (e.g., 54.3% versus 52.2% in serotype Gallinarum), the nucleotide content is not evenly distributed. For instance, regions of SPI-19 in serotype Gallinarum encoding Hcp-like proteins and proteins of unknown function presents a low G+C content (34.5%), while the rest of the island presents a G+C content significantly higher (59.1%).

SPI-19 encodes ~30 ORFs, including each T6SS core component (Figure 1 and Additional file 2). Notably, the island includes two ORFs encoding Hcp-like proteins that only share 27% of sequence identity (e.g., SG1029 and SG1043 in serotype Gallinarum). We refer to these proteins as Hcp-1 and Hcp-2, respectively (Figure 1 and Additional file 2). Hcp-1 is encoded in a putative operon that includes most of the T6SS functions. On the other hand, Hcp-2 is encoded upstream of a VgrG homolog at the opposite end of the island. This VgrG protein does not present a C-terminal extension or a putative effector domain. Additionally, the island includes 2 distantly related ImpA homologs (e.g., SG1030 and SG1032 in serotype Gallinarum), as reported for other T6SS loci [18]. SPI-19 does not encode transcriptional and post-transcriptional regulators present in others T6SS gene clusters [11,20], but encodes a protein containing the forkhead-associated (FHA) domain. It is known that an FHA-domain containing protein interacts with a Ser/Thr kinase (PpkA) and a phosphatase (PppA) to regulate type VI secretion in *P. aeruginosa* [11]. Although SPI-19 does not encode kinases and phosphatases, the possible interaction of the FHA-domain containing protein with other kinases and phosphatases to regulate secretion by this T6SS cannot be ruled out. In addition to T6SS-related components, a small "Sel1-like repeat protein" was identified in SPI-19 (Additional file 2). Serotypes Agona and Weltevreden encode a full version of the protein (SeAg_B1116 and SeW_A1263, respectively), while serotypes Enteritidis, Gallinarum and Dublin encode different truncated forms of it (SEN1008/SEN1009, SG1050/SG1051 and SeD_A1237, respectively). Notably, SeAg_B1116 and SeW_A1263 share ~31–38% sequence identity with sev-

**Figure 1**

Gene organization of T6SS gene clusters in *Salmonella*. Schematic representation of the T6SS loci identified in *Salmonella*, including the previously described T6SS encoded in SPI-6. One representative of each T6SS-encoding island is shown. ORFs are represented as blocked arrows showing the direction of their transcription. Conserved core, accessory and additional components are represented with a different color.

eral secreted Sel1-like proteins of *L. pneumophila* involved in manipulation of vacuolar trafficking in macrophages [23].

The genetic architecture of SPI-19 is highly conserved in serotypes Dublin, Weltevreden, Agona and Gallinarum. The main structural differences observed are circumscribed to the right end of the island, a region encoding a variable number of proteins with unknown function flanked by Rhs elements (Figure 2). Gene content variability in this region can be explained by deletions and rearrangements involving recombination events between Rhs elements. In serotype Gallinarum we identified a small ORF downstream of SG1040 that is absent in the genome annotation. This ORF encodes a gp25-like protein homolog to the T6SS-related ORF VCA0109 in *V. cholerae* (Additional file 2), and is annotated in the genome of the other SPI-19 positive serotypes.

In serotype Agona, the cluster of genes related to sugar metabolism flanking the left end of SPI-19 is missing. The absence of these genes has been reported previously [24]. The existence of 3 ORFs (SeAg_B1087 to SeAg_B1089)

encoding transposase remnants at this junction may be related to the absence of these genes. In addition, SPI-19 in serotype Agona includes a third ImpA homolog (SeAg_B1094) adjacent to a truncated second copy of IcmF (SeAg_B1095). Probably, these ORFs were generated by a duplication of neighboring genes encoding complete versions of both ImpA and IcmF proteins (SeAg_B1098 and SeAg_B1099, respectively).

In serotype Enteritidis, SPI-19 presents an internal deletion of ~24 kb with respect to the island in serotypes Dublin, Weltevreden, Agona and Gallinarum (Figure 2). As the result of this deletion, the island in Enteritidis only encodes 16 ORFs, 3 of which correspond to T6SS core components: SEN1002 (a putative Hcp-1 protein), SEN1003 (an ImpA homolog) and SEN1004 (a truncated form of IcmF). The absence of essential components suggests that the T6SS encoded in this serotype is not functional. Interestingly, the internal region of SPI-19 absent in Enteritidis has been defined as the "region of difference 9" (ROD9) in a comparative genome analysis of serotypes Gallinarum and Enteritidis [25]. It is tempting to speculate that the presence of an active T6SS is somehow related

Table 1: Distribution of T6SS loci in *Salmonella*

Genus	Species	Subspecies	Serotype	T6SS locus			
				SPI-6	SPI-19	SPI-20	SPI-21
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Typhi	X			
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Typhimurium	X			
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Paratyphi A	X			
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Paratyphi C	X			
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Kentucky	X			
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Newport	X			
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Heidelberg	X			
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Saintpaul	X ^a			
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Schwarzengrund	X			
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Tennessee	X			
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	4, [5] ,12:i-	X			
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Hadar	X			
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Infantis	X			
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Choleraesuis	X			
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Weltevreden	X	X		
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Dublin	X	X		
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Gallinarum		X		
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Agona		X		
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Enteritidis		X ^b		
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Paratyphi B				
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Virchow				
<i>Salmonella</i>	<i>enterica</i>	<i>enterica</i> (I)	Javiana				
<i>Salmonella</i>	<i>enterica</i>	<i>arizonae</i> (IIIa)	62:z4,z23:-			X	X
<i>Salmonella</i>	<i>enterica</i>	<i>diarizonae</i> (IIIb)	61:l,v:1,5,(7)				

^a SPI-6 in strain SARA29 presents an internal deletion of ~15 kb. In contrast, strain SARA23 harbors a full version of the island.

^b SPI-19 in serotype Enteritidis presents an internal deletion of ~24 kb.

S. Agona SL483

S. Dublin CT_02021853

S. Weltevreden SL484

S. Gallinarum 287/91

S. Enteritidis PT4 P125109

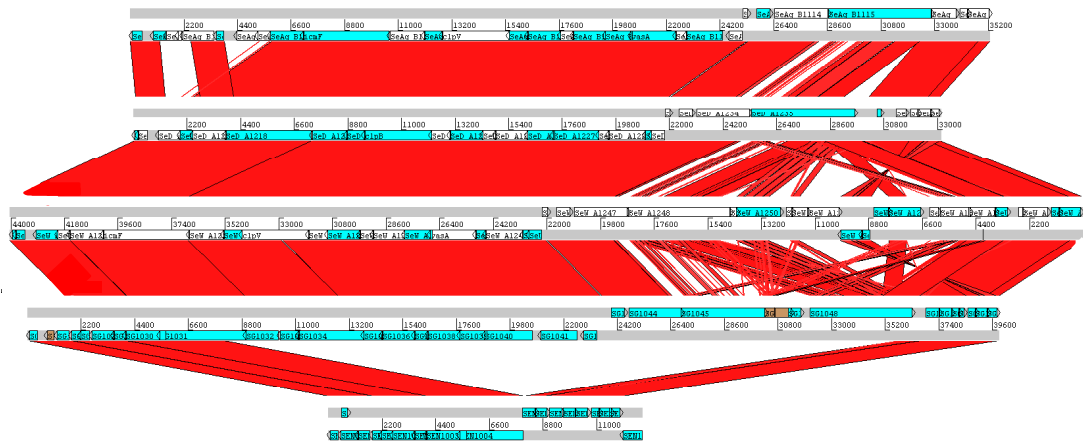


Figure 2
Comparative analysis of SPI-19 in *Salmonella*. DNA-based comparison of SPI-19 in serotypes Agona strain SL483, Dublin strain CT_02021853, Weltevreden strain SL484, Gallinarum strain 287/91 and Enteritidis strain P125109. BLASTN analysis was performed using WebACT and displayed with the ACT software.

to differences in host adaptation presented by these serotypes.

As mentioned before, SPI-19 is not obviously linked to a tRNA-encoding gene. Comparative genomic analyses revealed that SPI-19 is located at the left junction of a large chromosomal region (~900 to 960 kb, depending on the serotype) inverted with respect to the genome of serotype Typhimurium [25]. The right end of the inverted region is flanked by a functional copy of the tRNA-encoding gene *serX*. A more detailed sequence analysis revealed the presence of a 109 bp-long scar located at the left border of SPI-19 in the genomes of serotypes Dublin, Weltevreden, Gallinarum and Enteritidis. In serotype Agona, the last 48 nucleotides of the scar are lost. Notably, in each case the scar includes 15 nt in the 3' end of *serX*. Thus, a truncated version of *serX* is flanking the left end of SPI-19. These observations suggest that SPI-19 was originally inserted at the 3' end of *serX*, being flanked by a truncated and a functional copy of this tRNA-encoding gene. A posterior inversion event repositioned the functional copy of *serX* to a location distant from SPI-19 in the genome of serotypes Dublin, Weltevreden, Agona, Gallinarum and Enteritidis. It has been reported that insertions of new horizontally acquired genetic material could trigger compensating genomic rearrangements in order to maintain genome stability [26]. Thus, the acquisition of SPI-19 could be one of the events leading to the observed inversion. These events must have occurred after the divergence of serotype Typhimurium and a common ancestor of serotypes encoding SPI-19.

Subspecies *arizonae* (IIIa) encodes two novel T6SS gene clusters

Besides SPI-6 T6SS and SPI-19, two additional T6SS loci were identified in *Salmonella*. The novel gene clusters, now designated SPI-20 and SPI-21, are present only in the genome of *S. enterica* subspecies *arizonae* (IIIa) serotype 62:z4,z23:-. Notably, SPI-6 and SPI-19 are not encoded in the genome of this serotype.

SPI-20 corresponds to a ~34 kb genomic island encoding 28 ORFs (SARI_02707 to SARI_02736, coordinates 2617346-2651626) with an average G+C content of 53.1%. Seventeen ORFs related to T6SS function were identified, 13 of which correspond to core components (Figure 1 and Additional file 3). Among the T6SS-related components, the island encodes an Hcp-like protein (SARI_02729) and a VgrG protein with no C-terminal extension (SARI_02724). In addition, 3 ImpA homologs were identified: SARI_02718, SARI_02719 and SARI_02709 (Additional file 3). Sequence analyses revealed that SARI_02718 and SARI_02719 encode truncated proteins corresponding to the C-terminal and N-terminal of an ImpA homolog, respectively. Most probably,

both ORFs were generated as the result of a point mutation that introduced a stop codon in SARI_02719. On the other hand, SARI_02709 encodes an ImpA homolog distantly related to SARI_02719/SARI_02718. SPI-20 encodes neither transcriptional/post-transcriptional regulators nor FHA-domain containing proteins present in other T6SS loci [11,20]. As in the case of SPI-19, the G+C content of SPI-20 is not uniformly distributed. While most of the island has a G+C content of 55–60%, regions with low G+C content (~42%) encoding proteins with unknown function were detected. Although SPI-20 is located in the proximity of the tRNA-encoding gene *aspV* (Figure 3), the island is not flanked by obvious direct repeats. Interestingly, *aspV* corresponds to the insertion site of SPI-6 in many serotypes of *S. enterica*, indicating that different T6SS have been acquired at the same genomic location during *Salmonella* evolution.

SPI-21 is a second island encoding a T6SS in the genome of serotype IIIa 62:z4,z23:-. This island has an extension of ~55 kb and is delimited by transposase remnants (Additional file 4). Again, the average G+C content of the island (49.6%) is unevenly distributed, presenting regions with low G+C content (41.3%) encoding hypothetical proteins of unknown function and regions with higher G+C content (54.3%) encoding T6SS-related ORFs. SPI-21 is flanked by 2 horizontally acquired genetic elements (Figure 3). Adjacent to its right junction, a cryptic lambda-doid phage (SARI_02636 to SARI_02688) is inserted at the 3' end of *thrW*. This tRNA-encoding gene corresponds to the insertion site of P22 and related phages in *Salmonella* [27]. In addition, a small genomic island encoding 8 ORFs (SARI_2570 to SARI_02577) was found in the vicinity of the left junction of SPI-21. Five of these ORFs share extensive homology to ORFs in SPI-14 (STM0855 to STM0859 in serotype Typhimurium), initially described as a locus encoding genes required for full virulence of serotype Gallinarum in chickens [28].

SPI-21 includes 57 ORFs (SARI_02578 to SARI_02635, coordinates 2504768-2560524), 20 of which encode T6SS-related components (Figure 1 and Additional file 4). Among the ORFs not linked to T6SS, a putative chitinase (SARI_02620) and a protein (SARI_02608) containing the PASTA (penicillin-binding protein and Ser/Thr kinase associated) domain [29] were identified. Notably, SPI-21 presents additional ORFs linked to T6SS besides the core components shared with SPI-20. Hence, the island encodes not one but two VgrG proteins without C-terminal extension (SARI_02599 and SARI_02626). In addition, SPI-21 encodes an evolved VgrG protein (SARI_02603) containing a C-terminal extension that presents two conserved domains: COG5529 (S-type pyocin family) and PF01844 (HNH endonucleases) (Figure 4A). This constitutes one of the main findings of the

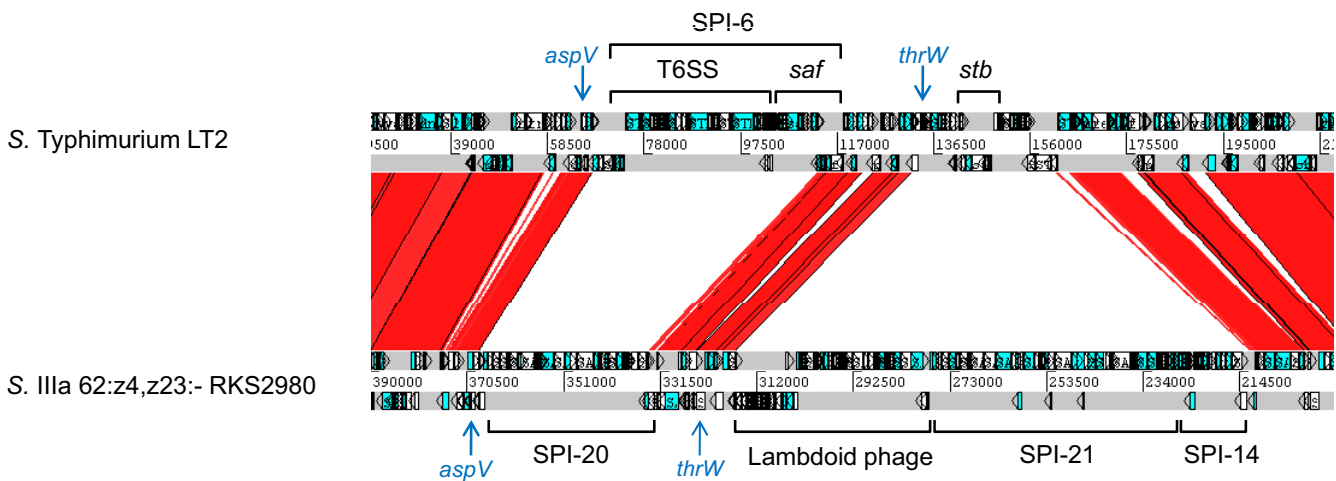


Figure 3

Genomic context of SPI-20 and SPI-21 in *S. IIIa 62:z4,z23:-*. DNA-based comparison of the genomic surrounding of SPI-20 and SPI-21 in *S. IIIa 62:z4,z23:-* strain RKS2980 and the corresponding region in the genome of *S. Typhimurium* strain LT2. BLASTN analysis was performed using WebACT and displayed with the ACT software. The location of key genetic elements is indicated. tRNA-encoding genes are indicated by blue arrows. *saf* and *stb* correspond to *Salmonella* fimbrial operons. SPI: *Salmonella* Pathogenicity Island.

present study as this is not only the first evolved VgrG described in *Salmonella*, but also the first evolved VgrG including a pyocin domain described so far in the literature. The C-terminal extension in SARI_02603 shares extensive sequence identity with the uropathogenic specific protein (Usp), a putative S-type pyocin encoded by UPEC strains [30] (Figure 4B). S-type pyocins are protease-sensitive, colicin-like bacteriocins produced by *P. aeruginosa* strains. These bacteriocins are normally constituted by two proteins: a large component that carries the killing activity and a small component that serves as an immunity protein [31]. Noteworthy, SPI-21 includes 4 ORFs encoding putative colicin/pyocin immunity proteins (SARI_02581, SARI_02585, SARI_02586 and SARI_02602), one of which is encoded just downstream of SARI_02603 (Additional file 4). Altogether, the presence of putative bacteriocin immunity proteins and an evolved VgrG with potential bacteriocin activity in SPI-21 suggests for the first time that bacteria could use a T6SS to kill other bacteria.

The SPI-6 T6SS gene cluster is widely distributed in *Salmonella enterica*

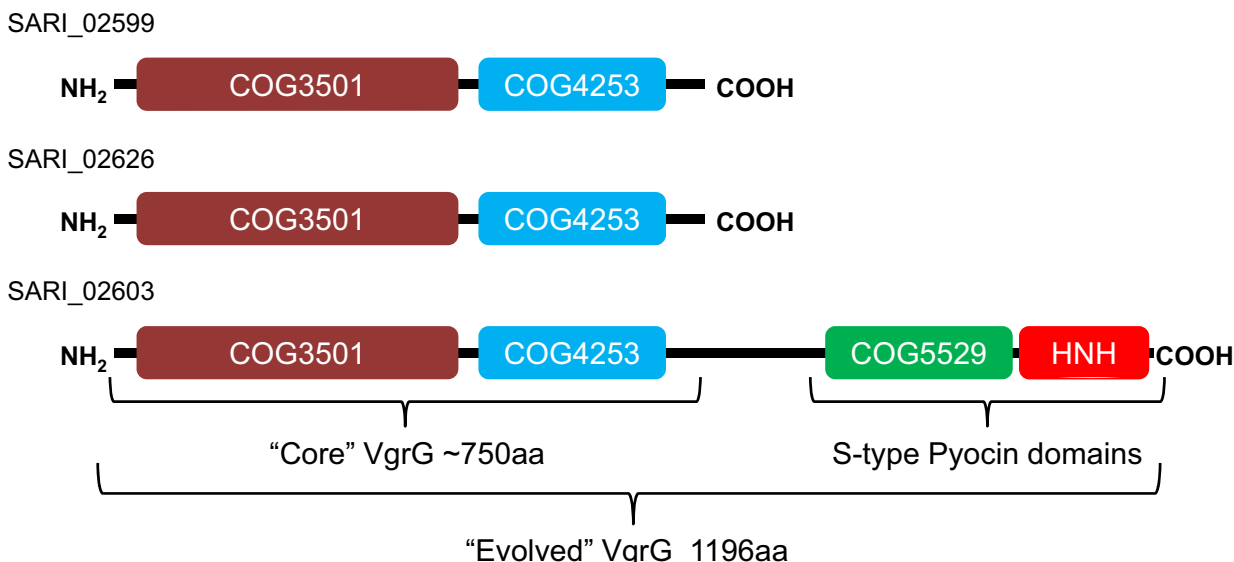
As mentioned before, the only T6SS identified to date in *Salmonella* is located in SPI-6, a pathogenicity island that also encodes the Saf and Tcf fimbrial systems [19,32]. The presence of the T6SS associated to SPI-6 was confirmed in 16 of the genomes surveyed, all of them belonging to *S. enterica* subspecies *enterica* (I). This group includes serotypes Typhimurium, Typhi, Paratyphi A, Paratyphi C, Kentucky, Newport, Heidelberg, Saintpaul, Schwarzengr-

und, Tennessee, Hadar, Infantis, Choleraesuis, Weltevreden, Dublin and I 4, [5], 12:i:- (Table 1). The ORFs, product sizes, conserved protein domains and genome coordinates of SPI-6 T6SS genes in representative strains of the mentioned serotypes are detailed in Additional file 5. Furthermore, the presence of the *saf* fimbrial operon was also detected in each of these serotypes, and the additional presence of the *tcf* fimbrial operon was confirmed in serotypes Paratyphi A, Choleraesuis and Typhi. This indicates that the 16 serotypes encoding the T6SS within SPI-6 harbor a complete version of this pathogenicity island.

Depending on the serotype, SPI-6 T6SS corresponds to a region of ~35 to 50 kb that encodes ~30 to 45 ORFs, including each of the 13 T6SS core components (Figure 1 and Additional file 5). Noteworthy, critical T6SS components in many SPI-6 T6SS are encoded by pseudogenes, as the case of *SciI* (*VipB* homolog) in serotype Typhi, and *SciG* (*ClpV* homolog) in serotypes I 4,[5],12:i:- and Choleraesuis (Additional file 5). This observation suggests that these T6SS are not functional.

The genetic architecture of SPI-6 T6SS is highly conserved among serotypes (Figure 5 and Additional file 5). However, structural differences limited to 3 particular regions in the island were observed. The first region is located downstream of *sciI* (*vipB* homolog), and encodes a variable number of proteins with no known function and embedded Hcp-like proteins. The obvious consequence of such variability is the presence of either one or two copies

A



B

```

SARI_02603      1  AFPAQATAAPLALCAATGLMTQVWGWBSLWALGARGTPEYVGAMMSALYVPSAGEGSDRVPGRDEFWVEEELRQKSLTGA
UspI           1  QPAEAAQTMLLFFGSV--LAAQIWGRWLSGLLSATCGS-VIIGALASALYIPPAGEGSGARVPGRDEFWVEEELRQKALAGS
consensus      1  *** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
SARI_02603     81  KATTRVRFWRQDEQCNMRYVGVHTGEGTPYEGVRTAEMKQWVVKNSRYEYFKPAHGADVPLITWTTPERPEGGDLPAHTGSD
UspI           78  TATTRVRFWGTDIHGKPOVYGVTGEGTPEYENRVRVANMOWNEQTORYEETPAHDVDPGLITWTTPENPEHGVVPGHTGND
consensus      81  * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
SARI_02603     161 IKPIDQATILVTPIDGKDEVTTPPFPMPEERDLNDYLLVFPAGSGIKPVVYVKDD
UspI           158 RPELEQPTILVTPIDGTDVYTPPFVVPDPKEFNDYLLVFPAGSGIKPTVYVKED
consensus      161 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
SARI_02603     241 LSVHNKNGAPTPSHIADKLGRS YGSFDLREDLWEEHISNDFKLMAQFGEDNOERTSNGL
UspI           238 MSVSNNGAPTPAHTADKLGRREFKTFDFREALWLEVSQDFELTAQFSSGNOETRIKQGI
consensus      241 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
SARI_02603     321 FQIHHDQAIEdGGGVYDLDNLRIVTPrLHDEIHyrR
UspI           318 FQIHRVAVEVGGSVYDIDNLRIVTPrLHDEIHyrR
consensus      321 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
  
```

Figure 4

VgrG homologs encoded in SPI-21. (A) Schematic representation of three VgrG proteins encoded in SPI-21. Conserved protein domains are represented in color. (B) Homology between the C-terminal extension of SARI_02603 and Usp in UPEC strains. SARI_02603 [residues 841–1201 (1201 total)] and Usp [residues 241–593 (593 total)] were aligned using the multiple sequence alignment tool ClustalW2. Identical and conserved residues are indicated by shaded boxes. Consensus residues are indicated below alignment with an asterisk. S-type pyocin and HNH domains are indicated in green and red color boxes, respectively.

of Hcp-like proteins in different serotypes. The second region is located downstream of *sciS* (*icmF* homolog) and encodes a variable number of proteins with unknown function. The third region is located downstream of *virgS* (*vgrG* homolog) and encodes a variable number of Rhs elements. In contrast to SPI-19, the variability observed in this region is not explained by deletions involving recombination between Rhs elements. More likely, different serotypes have acquired a distinct repertoire of unrelated Rhs-like elements.

Differences in the genetic structure of SPI-6 T6SS among strains of the same serotype were only detected in the case of *S. Saintpaul*. In this serotype, strain SARA29 presents a ~15 kb deletion in SPI-6 with respect to strain SARA23 (Figure 5). As the result of this deletion, SPI-6 T6SS in SARA29 does not encode critical components such as SciS (IcmF homolog), SciI (VipB homolog) and the two Hcp-like proteins, suggesting that the only T6SS encoded in this strain is not functional.

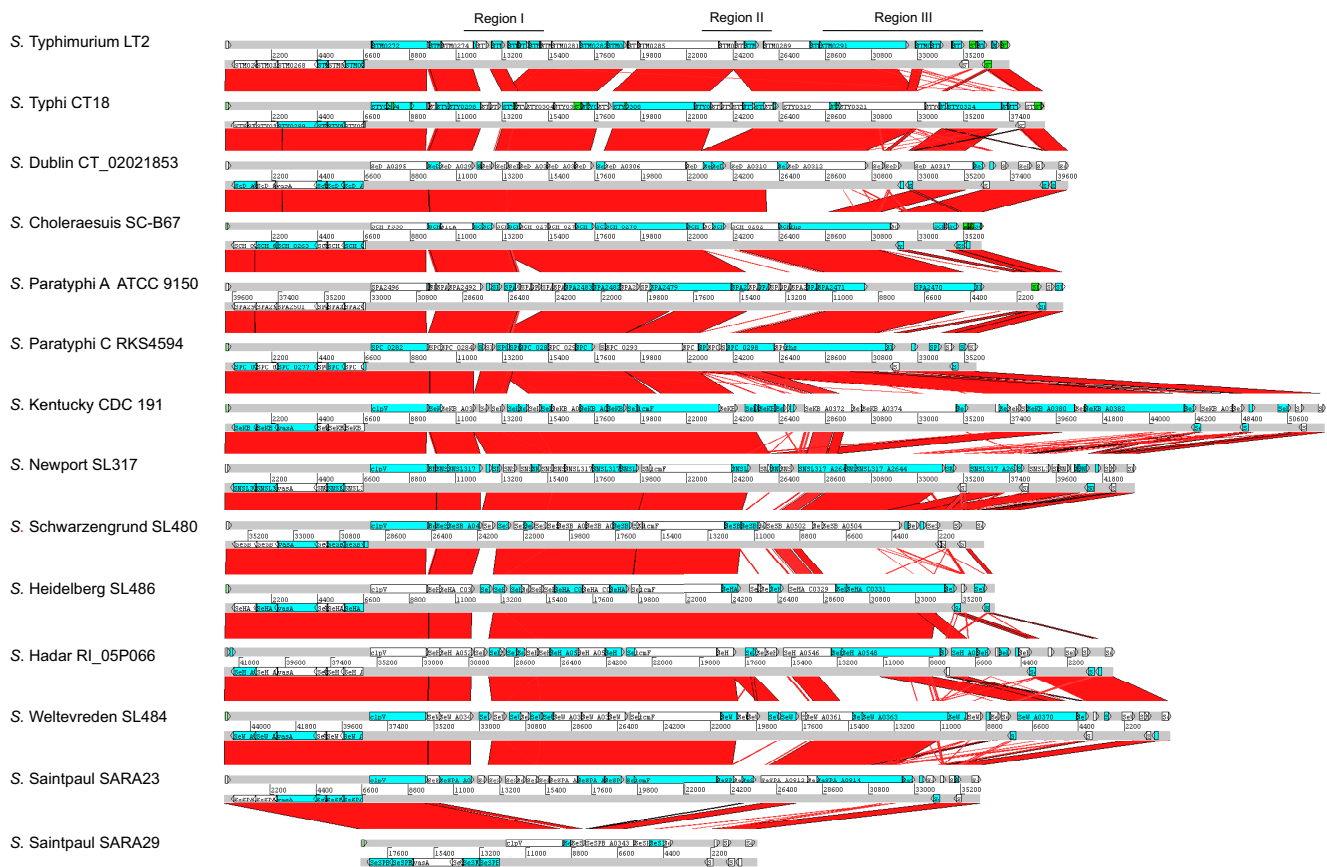


Figure 5
Genetic architecture of SPI-6 T6SS in *Salmonella enterica* serotypes. DNA-based comparison of SPI-6 T6SS in 13 different serotypes of *S. enterica* was performed by BLASTN analysis with WebACT and visualized with ACT software. The three major regions presenting structural differences between serotypes are indicated.

The analysis also revealed that SPI-6 in serotypes Enteritidis, Gallinarum, Agona, Paratyphi B and Virchow does not encode a T6SS (Table 1), but maintain the *saf* fimbrial operon. Furthermore, SPI-6 in serotype Javiana neither encodes a T6SS nor a fimbrial operon. These observations support the notion that the T6SS and the associated fimbrial operons *saf* and *tf* have been acquired independently during the evolution of *Salmonella* to produce a complex SPI-6. A more detailed sequence analysis revealed that SPI-6 in serotype Javiana only includes 5 ORFs, one of which corresponds to an ImpA-like protein associated to the SPI-6 T6SS. Several additional remnants of T6SS-related ORFs associated with SPI-6 were identified in serotypes Enteritidis, Gallinarum, Agona, Paratyphi B and Virchow (Additional file 6). The presence of these remnant ORFs suggests that an ancestor of SPI-6 T6SS was originally present in the genome of these serotypes, and was subsequently lost during evolution. In addition, a remnant of the first ORF in SPI-20 (and not linked to the T6SS function) was identified within SPI-6 in serotypes Gallinarum, Paratyphi B and Virchow (Additional file 6), suggesting that an ancestor of SPI-20 may have been

present at this location. This hypothesis is supported by the fact that the insertion site for both SPI-6 and SPI-20 in *Salmonella* corresponds to the same tRNA-encoding gene, *aspV*.

Organization of genes encoding core components in T6SS loci

Boyer and coworkers identified three groups of T6SS core components presenting a conserved genomic organization [8]. These conserved "modules" were defined by computing the number of loci in which every pair of T6SS core components (identified by their distinct COG ID) could be co-transcribed. The first module includes COG3521 (putative lipoprotein), COG3522, COG3455 (DotU) and COG3523 (IcmF). The second module includes COG3516 (VipA), COG3517 (VipB) and COG3157 (Hcp). The third module includes COG3518 (gp25-like protein), COG3519 (VasA) and COG3520 [8]. When we analyzed the conservation of these modules for each T6SS locus in *Salmonella*, important differences were identified:

- In the first module, components COG3521, COG3522 and COG3455 are linked in SPI-6 and SPI-19, while COG3523 is disconnected. In contrast, only COG3522 and COG3455 are linked in SPI-20 and SPI-21, while both COG3521 and COG3523 are disconnected. Furthermore, SPI-21 includes a second copy of COG3521 that is also unlinked to the proposed module.
- In the second module, COG3516 and COG3517 are linked in the 4 *Salmonella* T6SS loci, while COG3157 (Hcp) is not linked to either COG3516 or COG3517. In addition, both SPI-6 T6SS and SPI-19 encodes two Hcp proteins in different regions of each T6SS locus. Furthermore, in SPI-19 one of these ORFs (Hcp-1) is located adjacent to COG3516, but in a divergent orientation, clearly indicating that COG3516/COG3517 and Hcp have evolved independently. This is consistent with the report of Hcp proteins involved in T6SS function that are encoded outside the T6SS locus in *V. cholerae* [20], and also to the presence of "orphan Hcps" apparently unlinked to the T6SS function in *Salmonella* and other bacteria (see below).
- Finally, a conserved third module is present in SPI-6 T6SS and SPI-19. On the other hand, while COG3519 and COG3520 are linked in SPI-20 and SPI-21, COG3518 is disconnected from the module. Furthermore, SPI-21 includes 2 additional copies of COG3518 that are also disconnected from the module, and one from each other.

The similarities detected between modules in SPI-20 and SPI-21 indicate a close phylogenetic relationship, while differences between modules in SPI-6 T6SS, SPI-19 and SPI20/SPI-21 suggest that these loci have been acquired through independent horizontal transfer events, or have evolved independently to perform specialized functions.

Comparative analysis of *Salmonella* T6SS gene clusters

A basic comparative analysis of the *Salmonella* T6SS loci described in this study revealed that they share very limited homology one to each other at the DNA level (Additional file 7). This result supports the notion that these genomic islands were acquired by independent horizontal gene transfer events during *Salmonella* evolution. To identify evolutionary relationships between these T6SS gene clusters, we followed the approach used by Bingle and coworkers to study phylogeny between T6SS loci [1]. The concatenated aminoacidic sequences of conserved components VipA and VipB from 66 T6SS loci (including the original 37 loci used by Bingle and coworkers [1] and 22 loci representing every T6SS in *Salmonella*) were analyzed.

Notably, the T6SS loci in *Salmonella* differentially clustered in the mayor phylogenetic groups described for T6SS gene clusters [1] (Figure 6). Thus, every SPI-6 T6SS locus included in our analysis belongs to Group A, also including the T6SS gene clusters *h16_A0645-h16_A0657* in *Ralstonia eutropha* strain H16 and *tss-2* in *Burkholderia pseudomallei* strain 1106a. This result agrees with previous reports on the evolutionary relationship between SPI-6 T6SS in *Typhimurium* and *tss-2* [1,8]. In addition, SPI-20 and SPI-21 belong to Group C, which includes the T6SS cluster within PAI-*metV* in the uropathogenic *E. coli* (UPEC) strain CFT073. Finally, each SPI-19 analyzed belongs to Group D, also including the T6SS cluster within OI#7 in *E. coli* O157:H7 strain Sakai, and the AGI-1 island in the avian pathogenic *E. coli* (APEC) strain O1 (Figure 6). According to the nomenclature established in a recent phylogenetic analysis performed by Boyer and coworkers [8], SPI-6 T6SS, SPI-19 and SPI-20/SPI-21 belong to sub-groups III, I and II, respectively.

In a second analysis, we determined identity at the protein level of each *Salmonella* T6SS gene cluster, as a whole, to the closest phylogenetic relatives identified by our previous analysis. Notably, SPI-6 T6SS presented extensive identity and shared a similar genetic architecture with T6SS gene clusters *tss-2* in *B. pseudomallei* and *h16_A0645-h16_A0657* in *R. eutropha* (Figure 7A). The same was true for SPI-19 and the whole T6SS gene cluster within OI#7 in EHEC and the AGI-1 island in APEC (Figure 7B). On the other hand, SPI-20 and SPI-21 present several differences with its relative T6SS within PAI-*metV* in UPEC (Figure 7C). Thus, although SPI-20 conserves the overall genetic architecture of the T6SS locus in UPEC, the island includes several additional ORFs encoding proteins of unknown function (Additional file 3). In addition, SPI-21 presents a genetic architecture similar to SPI-20 and to the T6SS gene cluster in PAI-*metV*. Nevertheless, the island includes additional embedded ORFs encoding multiple copies of VgrG and gp25-like proteins, putative transposases, a putative chitinase and colicin/pyocin immunity proteins (Additional file 4). In spite of this, the phylogenetic proximity of SPI-20 and SPI-21 is evident.

The results obtained from both phylogenetic analyses reached the same conclusions, validating the evaluation of divergence in the aminoacidic sequence of VipA and VipB homologs to determine phylogenetic relationships between different T6SS loci. Altogether, these results strongly suggest that SPI-6 T6SS, SPI-19 and SPI-20/SPI-21 were acquired by independent horizontal transfer events from several unrelated bacterial species during the evolution of *Salmonella*. This is supported by the following observations: (i) each cluster is encoded in or corresponds to a genomic island that includes every T6SS core component, (ii) SPI-6 T6SS and SPI-20 are located adja-

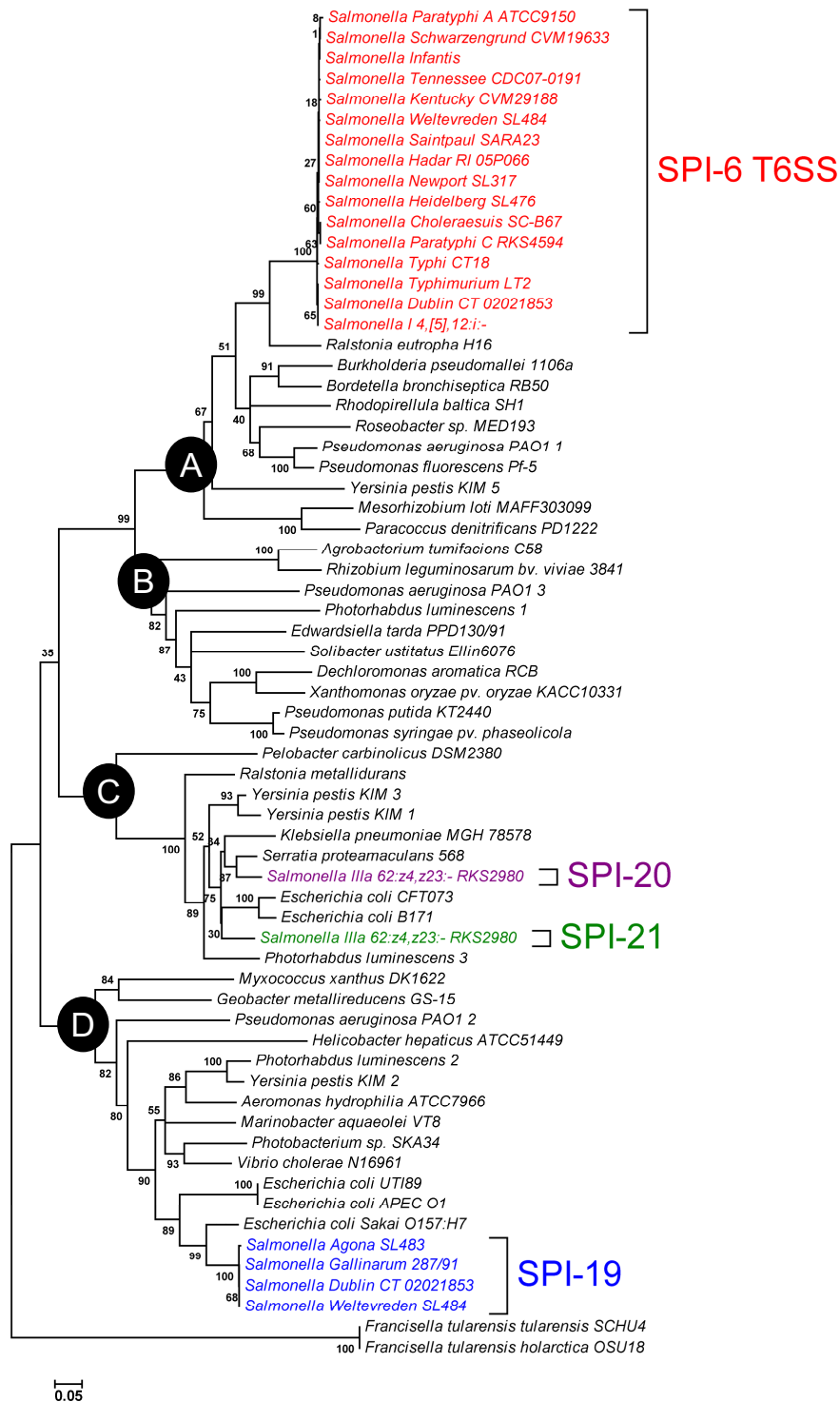


Figure 6
Evolutionary relationships of *Salmonella* T6SS loci. A distance tree (neighbour-joining) was calculated from concatenated VipA and VipB protein sequences of previously identified T6SS gene clusters, including the 3 novel *Salmonella* T6SS loci. Each of the four major phylogenetic groups is shown in the nodes labeled A to D. Bootstrap support values (% from 3,000 replicates) were: A, 99%; B, 80%; C, 99% and D, 99%.

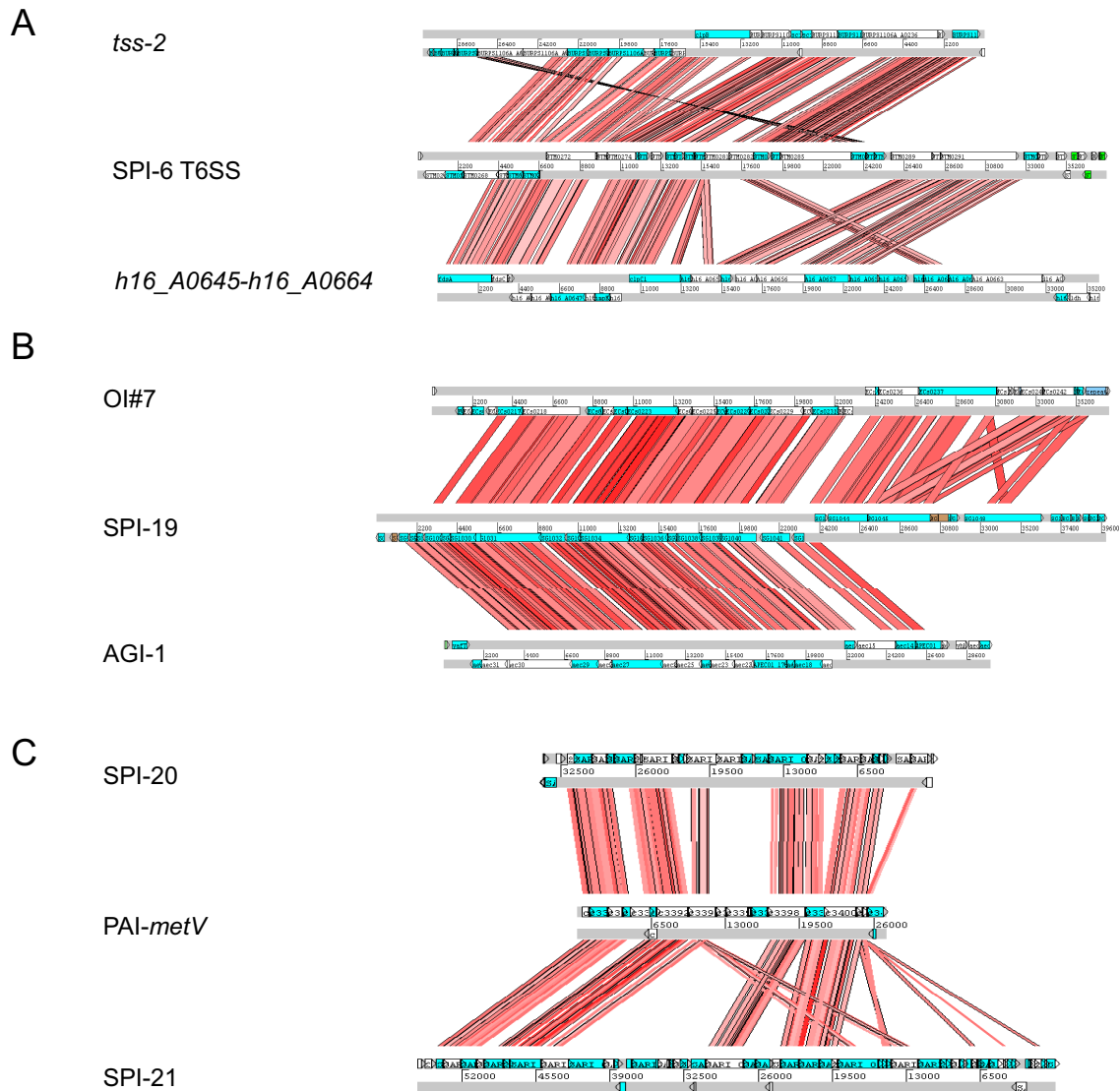


Figure 7
Comparative analysis of *Salmonella* T6SS clusters. DNA-based comparison of the T6SS encoded in SPI-6, SPI-19, SPI-20 and SPI-21 and phylogenetically-related T6SS loci. The analysis was performed by BLASTN with WebACT and visualized with ACT software. **(A)** Comparison of SPI-6 T6SS in serotype Typhimurium strain LT2 with gene cluster *tss-2* in *Burkholderia pseudomallei* strain 1106a and *h16_A0645-h16_A0664* in *Ralstonia eutropha* strain H16. **(B)** Comparison of SPI-19 in serotype Gallinarum strain 287/91 with the OI#7 island in *Escherichia coli* O157:H7 strain Sakai and island AGI-I in APEC strain O1. **(C)** Comparison of SPI-20 and SPI-21 in *S. IIIa* 62:z4,z23:- strain RKS2980 with the T6SS gene cluster within PAI-*metV* in UPEC strain CFT073.

cent to the tRNA-encoding gene *aspV*, a known hot-spot for insertion of genetic elements by horizontal transfer, (iii) SPI-19 is linked to a large chromosomal inversion, presumably caused by its insertion in the tRNA-encoding gene *serX*, and (iv) SPI-6 T6SS, SPI-19 and SPI-20/SPI-21 belong to different T6SS phylogenetic groups (Figure 6). Finally, the similar genetic organization presented by SPI-20 and SPI-21 (Figure 1), the presence of conserved regions sharing extensive sequence identity at the protein

level (Figure 7), and the fact that they belong to the same T6SS phylogenetic group (Figure 6) support a common origin for these T6SS loci. Thus, the presence of SPI-20 and SPI-21 in the genome of *S. IIIa* serotype 62:z4,z23:- can be explained either as the result of the acquisition of one of these clusters by lateral transfer and a posterior duplication event, or by sequential lateral transfer of each island from closely related donor organisms.

We can hypothesize evolutionary origins for each T6SS locus in *Salmonella* considering both phylogenetic analyses performed. Thus, SPI-6 T6SS may have been acquired from a beta-proteobacteria in the *Burkholderiaceae* family, related to *Burkholderia pseudomallei* and *Ralstonia eutropha* (recently reclassified as *Cupriavidus necator*). In contrast, SPI-19 seems to have been acquired from a pathogenic *E. coli* strain related to EHEC and APEC. Finally, SPI-20 and SPI-21 may have been acquired from a UPEC strain. In any case, each T6SS locus seems to have further evolved independently to generate systems performing specialized functions.

Orphan Hcp-like proteins in *Salmonella*

Several ORFs encoding additional Hcp-like proteins in *Salmonella* were identified in this study (Additional file 8). Because these proteins are not directly linked to T6SS gene clusters, we referred to them as "orphan Hcp-like proteins". According to their distribution, these proteins can be attributed to three major groups: (i) those present in every *S. enterica* serotype analyzed (e.g., STM4509.s and STM3131 in serotype Typhimurium), (ii) an unique representative present only in serotypes Agona, Paratyphi A, Kentucky and Typhi (e.g. SeAg_B3284 in serotype Agona), and (iii) those present only in *S. IIIa* serotype 62:z4,z23:- (i.e., SARI_00061, SARI_00912, SARI_01363 and SARI_03217). The genomic location and distinctive features of each of these proteins are detailed in Additional file 8.

Homologs of STM4509.s and STM3131 are widely distributed among *S. enterica* serotypes. Noteworthy, STM4509.s corresponds to Hile, a 148 aa protein that negatively regulates the expression of *hilA*, which encodes the master regulator controlling the expression of SPI-1 T3SS in *Salmonella* [33]. Folkesson and coworkers were the first to identify a similarity between Hile and Hcp proteins encoded in SPI-6 [19]. On the other hand, STM3131 corresponds to a 161 aa protein with no known function encoded in SPI-13, a locus including genes required for full virulence of serotypes Enteritidis, Gallinarum and Typhimurium in different animal hosts [28,34,35]. The genes encoding STM3131 and STM3130 constitute a small operon predicted to be part of the PmrAB regulon [36]. Remarkably, STM3130 belongs to a family of proteins including the baseplate gp25 protein in phage T4 and relatives of a conserved T6SS core component [8].

As mentioned, an orphan Hcp-like protein present only in serotypes Kentucky, Paratyphi A, Agona and Typhi was also identified. Depending on the serotype, this orphan Hcp corresponds to a protein of either 40 or 66 aa. In all cases, this small ORF is encoded in SPI-8 and is located downstream of an integrase remnant in the vicinity of the tRNA-encoding gene *pheV*. Finally, *S. IIIa* 62:z4,z23:- har-

bors four additional orphan Hcp-like proteins, each of them placed adjacent to transposase remnants or regions acquired by lateral gene transfer. Notably, SARI_03217 corresponds to an unusually long Hcp-like protein (403 aa versus ~160 aa in regular Hcp proteins) containing the Hcp/COG3157 domain at the N-terminus (150 aa) and a C-terminal extension (253 aa) including no conserved protein domain.

There are reports of other unusually long Hcp proteins carrying C-terminal extensions due to acquisition of novel protein domains. This is the case of Usp, a putative bacteriocin related to S-type pyocins containing an Hcp domain at the N-terminus [30]. To date, there is no experimental evidence confirming the putative bactericidal activity of Usp or explaining the role played by the Hcp domain in its function. However, the *usp* gene has been widely used as a virulence marker in UPEC strains [37-39]. Another example corresponds to the hypothetical protein YhhZ of *E. coli*. This protein also contains an Hcp domain at the N-terminus and a degenerate HNH domain at the C-terminus. The *yhhZ* gene has been reported to be induced during growth of *E. coli* in biofilms [40]. The structural features presented by these long Hcp proteins (i.e. an Hcp/COG3157 domain linked to an additional C-terminal domain) resemble those presented by the evolved VgrG proteins. Because of this similitude, we propose to designate these proteins as "evolved Hcps". Apparently, both Usp and YhhZ correspond to evolved Hcps not linked to T6SS function and seem to serve different roles for the bacterial cell. This could also be the case for SARI_03217 in *S. IIIa* 62:z4,z23:-.

In order to identify evolutionary relationships between these orphan Hcp-like proteins and those directly linked to T6SS loci, the sequence divergence among these proteins was analyzed. As a reference, 5 Hcp proteins described in *P. aeruginosa* strain PA01 (PA0263, PA1512, PA5267, PA2367 and PA0085) and 2 described in *V. cholerae* strain V52 (VCA_0017 and VC_1415) were included in the analysis. The results revealed the presence of 4 groups of Hcp relatives (Figure 8). Group A includes PA0085 and every Hcp protein encoded in SPI-6 T6SS. Group B includes PA2367 and every STM3131 homolog. Group C includes only Hile homologs, and Group D includes PA1512, PA5267, PA0263, VCA_0017, VC_1415, Hcp proteins in SPI-19, SPI-20, SPI-21, and the orphan Hcp-like proteins in *S. IIIa* 62:z4,z23:- (Figure 8).

Although evolutionary relationships between Hcp-like proteins were determined, the analysis failed to predict a role for many of them in T6SS function. We believe that this reflects the fact that the Hcp domain is linked to a particular folding pattern or structure rather than to a specific function or activity. Altogether, our data support the

notion that SARI_03217 and HilE homologs arose from a common ancestor of every Hcp protein and further evolved to gain new functions, while keeping structural properties of its relatives directly linked to T6SS function.

Conclusion

This study expands the current knowledge on bacterial type VI secretion by the identification and description of 3 novel T6SS loci in *Salmonella*. We have determined that each T6SS locus has a distinct evolutionary history, indicating acquisition through independent horizontal gene transfer events. Interestingly, while some serotypes harbor multiple T6SS loci others do not encode for any. This observation indicates that T6SS function is not essential for virulence but may provide an additional advantage in certain environments. On the other hand, the presence of 2 T6SS in many serotypes perhaps reflects specific roles for each of them in different steps during infection or in different environments or hosts. Our analysis also identified an evolved VgrG protein that contains a C-terminal extension sharing identity to S-type pyocins. To the best of our knowledge, this is the first description of a VgrG protein carrying a bacteriocin effector domain and suggests a novel role for T6SS in bacterial killing. The functionality of the 4 T6SS described here and their roles in *Salmonella* pathogenesis and host-specificity are currently under study in our laboratory.

Methods

Salmonella genomic data acquisition

Information about the current status of *Salmonella* genome sequencing projects was obtained from the

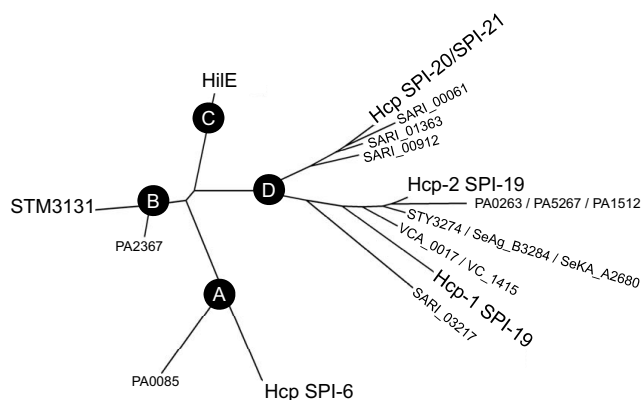


Figure 8
Evolutionary relationships of Hcp-like protein. A distance tree (neighbour-joining) was calculated from the alignment of every Hcp-like protein identified in *Salmonella*. The analysis also included Hcp proteins of *Pseudomonas aeruginosa* strain PA01 and *Vibrio cholerae* strain V52. Each of the major groups of Hcp phylogeny is shown in the nodes labeled A to D. Bootstrap support values (% from 3,000 replicates) were: A, 98%; B, 94%; C, 98% and D, 80%.

Genomes On Line Database (GOLD; updated on May 2, 2009) ([41], web site: <http://www.genomesonline.org/>) and from Table 1 in reference [42]. The 44 *Salmonella* sequencing projects having sequence data publicly available were selected for genome-wide *in silico* identification of T6SS loci. The genomes analyzed covered 24 different serotypes of *Salmonella enterica* (22 in subspecies I, 1 in subspecies IIIa and 1 in subspecies IIIb). Genome accession numbers and information about the *Salmonella* genome sequencing projects utilized in this work are detailed in the Additional file 1.

In silico identification of T6SS loci

Nucleotide sequence of ORFs representing T6SS core components were obtained from public sequence databases. Three representatives of each branch in T6SS phylogenetic trees previously described [1,8] (Additional file 9) were used as baits in sequential BLASTN, BLASTP and TBLASTX searches [43-45] to identify T6SS homologs in the 44 *Salmonella* genomic sequences listed in Additional file 1. When a core component was identified, a systematic analysis was performed to detect adjacent T6SS components. A T6SS locus was defined as a gene cluster encoding at least 5 core components. Boundaries of the genomic islands encoding T6SS were defined by thorough nucleotide sequence analysis. A systematic analysis of gene content and genetic architecture of the whole T6SS gene clusters identified was performed for 31 annotated and fully assembled genomes. The remaining 13 genomes correspond to unfinished projects; therefore in many cases an exhaustive ORF-by-ORF analysis of T6SS genetic architecture was not possible. In this case, we only determined the presence/absence status for T6SS core components in unassembled contigs.

DNA and protein sequence analysis

Nucleotide sequences were analyzed by the sequence visualization and annotation tool Artemis version 10 [46] and the Vector NTI Advanced software version 11.0 (Invitrogen). Graphical representations of T6SS gene clusters were generated with the CLC Sequence Viewer software version 6.0.2 (CLC bio). For comparative analysis, nucleotide sequences were aligned by BLASTN and TBLASTX with the WebACT online resource ([47], web site: <http://www.webact.org/>) and visualized with the Artemis Comparison Tool (ACT) release 6 [48]. Presence of conserved protein domains in each ORF encoded in the identified T6SS loci was determined by cross-reference of several publicly available databases: The Simple Modular Architecture Research Tool (SMART) ([49], web site: <http://smart.embl.de/>), the Conserved Domain Database and Search Service version 2.16 ([50], web site: <http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) and the Clusters of Orthologous Groups of proteins (COGs) database ([51], web site: <http://www.ncbi.nlm.nih.gov/COG>).

Phylogenetic analyses

Phylogenetic analyses were performed with the Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.2 [52] based on the approach described by Bingle and coworkers [1]. In the case of T6SS loci, the concatenated aminoacidic sequences of VipA and VipB homologs encoded in 66 T6SS loci (Additional file 10) were aligned using ClustalW [53] with the default parameters. A similar alignment was carried out using 113 Hcp-like proteins from strains representing each serotype of *Salmonella* studied, *Pseudomonas aeruginosa* strain PAO1 and *Vibrio cholerae* strain V52 (Additional file 8). Phylogenetic trees were built from the alignments by the bootstrap test of phylogeny (3,000 replications) using the neighbor-joining method with a Poisson correction model.

Authors' contributions

CB, CS and IC conceived the study. CB acquired the genomic sequences and performed most of the bioinformatic and phylogenetic analyses. CS carried out bioinformatic and phylogenetic analyses. JJ carried out bioinformatic analyses and helped in acquisition of genomic sequences. CB and CS interpreted the data generated. CB, CS and IC drafted the manuscript. CB and JJ prepared figures, tables and additional files presenting the data. All authors read and approved the final manuscript.

Additional material

Additional file 1

Salmonella genome sequencing projects utilized in this work. This file includes information on the 44 *Salmonella* genome sequencing projects used in this report.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-354-S1.xls>]

Additional file 2

Gene content of SPI-19 in different Salmonella serotypes. This file includes the genomic coordinates, product sizes and conserved protein domains of every ORF encoded in SPI-19 in Enteritidis strain P125109, Gallinarum strain 287/91, Agona strain SL483, Dublin strain CT_02021853 and Weltevreden strain SL484. In addition, each component is compared to the corresponding homolog in the OI#7 T6SS gene cluster of *Escherichia coli* O157:H7 strain Sakai.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-354-S2.xls>]

Additional file 3

Gene content of SPI-20 in Salmonella IIIa 62:z4,z23:-. This file includes the genomic coordinates, product sizes and conserved protein domains of every ORF encoded in SPI-20 in *Salmonella* IIIa 62:z4,z23:- strain RKS2980. In addition, each component is compared to the corresponding homolog in the T6SS gene cluster within PAI-metV in UPEC strain CFT073.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-354-S3.xls>]

Additional file 4

Gene content of SPI-21 in Salmonella IIIa 62:z4,z23:-. This file includes the genomic coordinates, product sizes and conserved protein domains of every ORF encoded in SPI-21 in *Salmonella* IIIa 62:z4,z23:- strain RKS2980. In addition, each component is compared to the corresponding homolog in the T6SS gene cluster within PAI-metV in UPEC strain CFT073.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-354-S4.xls>]

Additional file 5

Gene content of SPI-6 T6SS in different Salmonella serotypes. This file includes the genomic coordinates, product sizes and conserved protein domains of every ORF encoded in SPI-6 T6SS in different *Salmonella* enterica serotypes. In addition, each component is compared to the corresponding homolog in serotype Typhimurium strain LT2.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-354-S5.xls>]

Additional file 6

Presence of T6SS-related ORFs associated with SPI-6 T6SS and SPI-20 in Salmonella serotypes encoding remnants of these islands. This document presents information about ORFs related to SPI-6 T6SS and SPI-20 in serotypes that do not encode these T6SS gene clusters. Reference ORFs for SPI-6 T6SS and SPI-20 were obtained from *S. Typhimurium* strain LT2 and *S. IIIa 62:z4,z23:-* strain RKS2980, respectively.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-354-S6.xls>]

Additional file 7

Comparative analysis of T6SS gene clusters encoded in Salmonella. This figure shows a DNA-based comparison of the 4 T6SS loci of *Salmonella*. BLASTN analysis of one representative of each genomic island (SPI-6 T6SS, SPI-19, SPI-20 and SPI-21) was performed using WebACT and visualized with ACT software.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-354-S7.pdf>]

Additional file 8

Hcp-like proteins in Salmonella. This file includes a list of every Hcp-like protein in *Salmonella* and reference Hcp proteins in *Pseudomonas aeruginosa* and *Vibrio cholerae* used in phylogenetic analysis.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-354-S8.xls>]

Additional file 9

T6SS components used as baits in genome-wide screens for novel T6SS loci in Salmonella. This table includes a list of T6SS core components in different phylogenetic groups that were used in this study as baits to identify novel *Salmonella* T6SS loci by in silico analyses.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-354-S9.xls>]

Additional file 10

VipA/VipB homologs utilized to perform T6SS phylogenetic analysis. This file includes a list of *VipA/VipB* homologs in 66 T6SS loci used in this study to identify T6SS phylogeny.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-354-S10.xls>]

Acknowledgements

We thankfully acknowledge the following centers for providing genome sequencing data: The Wellcome Trust Sanger Institute for the sequences of *S. enterica* serotype Hadar, *S. enterica* serotype Infantis and *S. enterica* serotype Typhimurium strains DT104 and SL1344 (data available at <ftp://ftp.sanger.ac.uk/pub/pathogens/Salmonella/>). The Genome Sequencing Center at Washington University School of Medicine in St. Louis for the sequences of *S. enterica* serotype Paratyphi A strain ATCC 9150, *S. enterica* serotype Paratyphi B strain SPB7, *S. enterica* subspecies IIIa serotype 62:z4,z23:- strain RKS2980 and *S. enterica* subspecies IIIb serotype 61:1,v:1.5,(7) strain CDC 01-0005 (data available at ftp://genome.wustl.edu/pub/organism/Microbes/Enteric_Bacteria/). The Department of Microbiology at the University of Illinois for the sequence of *S. enterica* serotype Enteritidis strain LK5 (data available at <http://www.salmonella.org>).

We are indebted to Cecilia Toro and Mercedes Zaldivar for continuous support and insightful discussions. This work was supported by grant ADI-08/2006 to IC from "Programa Bicentenario de Ciencia y Tecnología" (CONICYT, Chile) and The World Bank. CJB was supported by a fellowship from CONICYT (Chile).

References

- Bingle LE, Bailey CM, Pallen MJ: **Type VI secretion: a beginner's guide.** *Curr Opin Microbiol* 2008, **11**(1):3-8.
- Cascales E: **The type VI secretion toolkit.** *EMBO Rep* 2008, **9**(8):735-741.
- Filloux A, Hachani A, Bleves S: **The bacterial type VI secretion machine: yet another player for protein transport across membranes.** *Microbiology* 2008, **154**(Pt 6):1570-1583.
- Pukatzki S, McAuley SB, Miyata ST: **The type VI secretion system: translocation of effectors and effector-domains.** *Curr Opin Microbiol* 2009, **12**(1):11-17.
- Das S, Chaudhuri K: **Identification of a unique IAHP (IcmF associated homologous proteins) cluster in *Vibrio cholerae* and other proteobacteria through in silico analysis.** *In Silico Biol* 2003, **3**(3):287-300.
- Sexton JA, Miller JL, Yoneda A, Kehl-Fie TE, Vogel JP: ***Legionella pneumophila* DotU and IcmF are required for stability of the Dot/Icm complex.** *Infect Immun* 2004, **72**(10):5983-5992.
- Zusman T, Feldman M, Halperin E, Segal G: **Characterization of the *icmH* and *icmF* genes required for *Legionella pneumophila* intracellular growth, genes that are present in many bacteria associated with eukaryotic cells.** *Infect Immun* 2004, **72**(6):3398-3409.
- Boyer F, Fichant G, Berthod J, Vandenbrouck Y, Attree I: **Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learnt from available microbial genomic resources?** *BMC Genomics* 2009, **10**(1):104.
- Mougous JD, Cuff ME, Raunser S, Shen A, Zhou M, Gifford CA, Goodman AL, Joachimiak G, Ordóñez CL, Lory S, et al.: **A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus.** *Science* 2006, **312**(5779):1526-1530.
- Zheng J, Leung KY: **Dissection of a type VI secretion system in *Edwardsiella tarda*.** *Mol Microbiol* 2007, **66**(5):1192-1206.
- Mougous JD, Gifford CA, Ramsdell TL, Mekalanos JJ: **Threonine phosphorylation post-translationally regulates protein secretion in *Pseudomonas aeruginosa*.** *Nat Cell Biol* 2007, **9**(7):797-803.
- Schlieker C, Zentgraf H, Dersch P, Mogk A: **ClpV, a unique Hsp100/Clp member of pathogenic proteobacteria.** *Biol Chem* 2005, **386**(11):1115-1127.
- Bonemann G, Pietrosiuk A, Diemand A, Zentgraf H, Mogk A: **Remodelling of *VipA/VipB* tubules by ClpV-mediated threading is crucial for type VI protein secretion.** *EMBO J* 2009, **28**(4):315-325.
- Leiman PG, Basler M, Ramagopal UA, Bonanno JB, Sauder JM, Pukatzki S, Burley SK, Almo SC, Mekalanos JJ: **Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin.** *Proc Natl Acad Sci USA* 2009, **106**(11):4154-4159.
- Pell LG, Kanelis V, Donaldson LW, Howell PL, Davidson AR: **The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system.** *Proc Natl Acad Sci USA* 2009, **106**(11):4160-4165.
- Pukatzki S, Ma AT, Revel AT, Sturtevant D, Mekalanos JJ: **Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin.** *Proc Natl Acad Sci USA* 2007, **104**(39):15508-15513.
- Suarez G, Sierra JC, Sha J, Wang S, Erova TE, Fadl AA, Foltz SM, Horneman AJ, Chopra AK: **Molecular characterization of a functional type VI secretion system from a clinical isolate of *Aeromonas hydrophila*.** *Microb Pathog* 2008, **44**(4):344-361.
- Shrivastava S, Mande SS: **Identification and functional characterization of gene components of Type VI Secretion system in bacterial genomes.** *PLoS ONE* 2008, **3**(8):e2955.
- Folkesson A, Lofdahl S, Normark S: **The *Salmonella enterica* subspecies I specific centisome 7 genomic island encodes novel protein families present in bacteria living in close contact with eukaryotic cells.** *Res Microbiol* 2002, **153**(8):537-545.
- Pukatzki S, Ma AT, Sturtevant D, Krastins B, Sarracino D, Nelson WC, Heidelberg JF, Mekalanos JJ: **Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system.** *Proc Natl Acad Sci USA* 2006, **103**(5):1528-1533.
- Parsons DA, Heffron F: ***sciS*, an *icmF* homolog in *Salmonella enterica* serovar Typhimurium, limits intracellular replication and decreases virulence.** *Infect Immun* 2005, **73**(7):4338-4345.
- Klumpp J, Fuchs TM: **Identification of novel genes in genomic islands that contribute to *Salmonella typhimurium* replication in macrophages.** *Microbiology* 2007, **153**(Pt 4):1207-1220.
- Newton HJ, Sansom FM, Dao J, McAlister AD, Sloan J, Cianciotto NP, Hartland EL: **Sell repeat protein LpnE is a *Legionella pneumophila* virulence determinant that influences vacuolar trafficking.** *Infect Immun* 2007, **75**(12):5575-5585.
- Reen FJ, Boyd EF, Porwollik S, Murphy BP, Gilroy D, Fanning S, McClelland M: **Genomic comparisons of *Salmonella enterica* serovar Dublin, Agona, and Typhimurium strains recently isolated from milk filters and bovine samples from Ireland, using a *Salmonella* microarray.** *Appl Environ Microbiol* 2005, **71**(3):1616-1625.
- Thomson NR, Clayton DJ, Windhorst D, Vernikos G, Davidson S, Churcher C, Quail MA, Stevens M, Jones MA, Watson M, et al.: **Comparative genome analysis of *Salmonella* Enteritidis PT4 and *Salmonella* Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways.** *Genome Res* 2008, **18**(10):1624-1637.
- Hughes D: **Evaluating genome dynamics: the constraints on rearrangements within bacterial genomes.** *Genome Biol* 2000, **1**(6):REVIEWS0006.
- Kropinski AM, Sulakvelidze A, Konczyk P, Poppe C: ***Salmonella* phages and prophages – genomics and practical aspects.** *Methods Mol Biol* 2007, **394**:133-175.
- Shah DH, Lee MJ, Park JH, Lee JH, Eo SK, Kwon JT, Chae JS: **Identification of *Salmonella gallinarum* virulence genes in a chicken infection model using PCR-based signature-tagged mutagenesis.** *Microbiology* 2005, **151**(Pt 12):3957-3968.
- Yeats C, Finn RD, Bateman A: **The PASTA domain: a beta-lactam-binding domain.** *Trends Biochem Sci* 2002, **27**(9):438.

30. Parret AH, De Mot R: **Escherichia coli's uropathogenic-specific protein: a bacteriocin promoting infectivity?** *Microbiology* 2002, **148(Pt 6)**:1604-1606.
31. Michel-Briand Y, Baysse C: **The pyocins of Pseudomonas aeruginosa.** *Biochimie* 2002, **84(5-6)**:499-510.
32. Folkesson A, Advani A, Sukupolvi S, Pfeifer JD, Normark S, Lofdahl S: **Multiple insertions of fimbrial operons correlate with the evolution of Salmonella serovars responsible for human disease.** *Mol Microbiol* 1999, **33(3)**:612-622.
33. Baxter MA, Fahlen TF, Wilson RL, Jones BD: **HilE interacts with HILD and negatively regulates hilA transcription and expression of the Salmonella enterica serovar Typhimurium invasive phenotype.** *Infect Immun* 2003, **71(3)**:1295-1305.
34. Zhao Y, Jansen R, Gastra W, Arkesteijn G, Zeijst BA van der, van Putten JP: **Identification of genes affecting Salmonella enterica serovar enteritidis infection of chicken macrophages.** *Infect Immun* 2002, **70(9)**:5319-5321.
35. Santiviago CA, Reynolds MM, Porwollik S, Choi SH, Long F, Andrews-Polymenis HL, McClelland M: **Analysis of pools of targeted Salmonella deletion mutants identifies novel genes affecting fitness during competitive infection in mice.** *PLoS Pathog* 2009, **5(7)**:e1000477.
36. Marchal K, De Keersmaecker S, Monsieurs P, van Boxel N, Lemmens K, Thijs G, Vanderleyden J, De Moor B: **In silico identification and experimental validation of PmrAB targets in Salmonella typhimurium by regulatory motif detection.** *Genome Biol* 2004, **5(2)**:R9.
37. Kanamaru S, Kurazono H, Ishitoya S, Terai A, Habuchi T, Nakano M, Ogawa O, Yamamoto S: **Distribution and genetic association of putative uropathogenic virulence factors iroN, iha, kpsMT, ompT and usp in Escherichia coli isolated from urinary tract infections in Japan.** *J Urol* 2003, **170(6 Pt 1)**:2490-2493.
38. Kurazono H, Nakano M, Yamamoto S, Ogawa O, Yuri K, Nakata K, Kimura M, Makino S, Nair GB: **Distribution of the usp gene in uropathogenic Escherichia coli isolated from companion animals and correlation with serotypes and size-variations of the pathogenicity island.** *Microbiol Immunol* 2003, **47(10)**:797-802.
39. Nakano M, Yamamoto S, Terai A, Ogawa O, Makino SI, Hayashi H, Nair GB, Kurazono H: **Structural and sequence diversity of the pathogenicity island of uropathogenic Escherichia coli which encodes the USP protein.** *FEMS Microbiol Lett* 2001, **205(1)**:71-76.
40. Domka J, Lee J, Bansal T, Wood TK: **Temporal gene-expression in Escherichia coli K-12 biofilms.** *Environ Microbiol* 2007, **9(2)**:332-346.
41. Liolios K, Mavromatis K, Tavernarakis N, Kyrpides NC: **The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata.** *Nucleic Acids Res* 2008:D475-479.
42. Andrews-Polymenis HL, Santiviago CA, McClelland M: **Novel genetic tools for studying food-borne Salmonella.** *Curr Opin Biotechnol* 2009, **20**:149-157.
43. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215(3)**:403-410.
44. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25(17)**:3389-3402.
45. States DJ, Gish W: **Combined use of sequence similarity and codon bias for coding region identification.** *J Comput Biol* 1994, **1(1)**:39-50.
46. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B: **Artemis: sequence visualization and annotation.** *Bioinformatics* 2000, **16(10)**:944-945.
47. Abbott JC, Aanensen DM, Bentley SD: **WebACT: an online genome comparison suite.** *Methods Mol Biol* 2007, **395**:57-74.
48. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J: **ACT: the Artemis Comparison Tool.** *Bioinformatics* 2005, **21(16)**:3422-3423.
49. Letunic I, Doerks T, Bork P: **SMART 6: recent updates and new developments.** *Nucleic Acids Res* 2009:D229-232.
50. Marchler-Bauer A, Anderson JB, Derbyshire MK, DeWeese-Scott C, Gonzales NR, Gwadz M, Hao L, He S, Hurwitz DI, Jackson JD, et al.: **CDD: a conserved domain database for interactive domain family analysis.** *Nucleic Acids Res* 2007:D237-240.
51. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, et al.: **The COG database: an updated version includes eukaryotes.** *BMC Bioinformatics* 2003, **4**:41.
52. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24(8)**:1596-1599.
53. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22(22)**:4673-4680.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

