

## SPI-7: *Salmonella*'s Vi-Encoding Pathogenicity Island

Helena M. B. Seth-Smith

Pathogen Sequencing Unit, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom

### Abstract

*Salmonella* Pathogenicity Island-7 (SPI-7) is a large, mosaic, genetic island, found in several serovars of *Salmonella enterica* subsp. *enterica* associated with systemic disease. As well as encoding genes which may aid its own transmission, it carries genes for potential virulence factors such as Vi antigen, SopE effector and type IVB pili. The stability of SPI-7 is of interest with respect to typhoid fever and related vaccines.

**Key Words:** SPI7, Typhi, typhoid, virulence, PAI, *Salmonella*

*J Infect Developing Countries* 2008; 2(4): 267-271.

Received 27 June 2008 - Accepted 22 July 2008

Copyright © 2008 Seth-Smith. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Introduction

*Salmonella* Pathogenicity Island-7 (SPI-7) is the largest genomic island yet identified in *Salmonella*, comprising up to 134 kb. It was first discovered as a large insertion in the genome of the human restricted pathogen *Salmonella enterica* subsp. *enterica* serovar Typhi (*S. Typhi*), relative to that of serovar Typhimurium (*S. Typhimurium*) [1]. The "major pathogenicity island" was renamed SPI-7 during the annotation of the *S. Typhi* strain CT18 genome, after the discovery of a number of other *S. Typhi*-specific islands [2]. The genome annotation revealed that SPI-7 incorporates approximately 150 predicted genes, systematically numbered STY4521-STY4680. SPI-7 is also carried by *Salmonella enterica* subsp. *enterica* serovar Paratyphi C (*S. Paratyphi C*) [3] and some strains of serovar Dublin (*S. Dublin*) [4,5]. There are some differences between the versions of the island in different strains and serovars, yet it is still recognised as SPI-7 [6] (Figure). Although tempting to correlate its presence with the ability of the strains to cause systemic disease in humans, serovar Paratyphi A (*S. Paratyphi A*) causes enteric disease in the absence of SPI-7 [7].

SPI-7 fulfils all the criteria necessary to label it a pathogenicity island (PAI) [8]: it is large, carries putative virulence genes, has been implicated in

enhancing pathogenicity and possesses a G+C content which differs from that of the genomic backbone. It is thought to integrate and excise as a conjugative transposon or, more likely, phage [9]. SPI-7 is sited at the tRNA<sup>PheU</sup> gene [10], between an intact copy of the tRNA and a partial duplicated copy [5]. The structure of the island is modular, and appears to have arisen through serial acquisition of functional units (Figure). The mosaic structure of SPI-7 comprises regions thought to be involved in the island's mobility, and several regions implicated in virulence: the locus for production and export of Vi antigen (*viaB*-locus), SopE phage and a type IVB pilus locus.

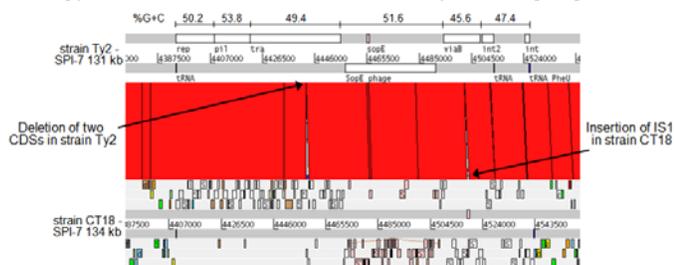
### Stability of SPI-7 and the *S. Typhi* genome

Very few clinical isolates of *S. Typhi* lack SPI-7 in their genomic backbone, implying a selective advantage in carrying the island. A survey of over 2,000 clinical isolates recently found only one strain which has lost SPI-7, by precise excision [11]. Loss appears to occur more frequently after storage, with *S. Typhi* strain Ty2 losing the majority of SPI-7 during passage, excising the island from STY4521-STY4666 [9]. This deletion derivative was, interestingly, found to be more invasive of human epithelial cells than the parental strain with the complete SPI-7 [9]. In addition, 8 out of 120 stored clinical isolates of *S. Typhi* were

found to have lost SPI-7 either through precise excision or imprecise removal of the region [12], and *S. Typhi* strain SARB64 has been found, by microarray, to lack all SPI-7 associated genes [13]. This implies that although SPI-7 can be lost from strains, it is generally maintained by some selective pressure in clinical settings.

**Figure.** Artemis Comparison Tool (ACT) [53] representation comparing the SPI-7 regions of the genomes of two *S. Typhi* strains: strain Ty2 (AE014613 - top) and strain CT18 (AL513382 - bottom).

Red bars indicate regions of >99% nucleotide identity. The numbers refer to coordinates in the published genomes. The G+C% of the regions is displayed, demonstrating the mosaic nature of the island. The modular structure is illustrated by white bars on the strain Ty2 sequence, with the predicted coding sequences (CDSs) shown as coloured blocks on the strain CT18 sequence in each of the six reading frames: *rep* - homology to plasmid replication region, STY4521-4537; *pil* - type IV pilus operon, STY4539-4553; *tra* - homology to plasmid transfer region, STY4554-4595; *sopE* - SopE phage, STY4600-4645; *viaB*-locus - *viaB*-locus operon, STY4651-4662; *int2* - integrase region, STY4664-4666; *int* - integrase, STY4680. The tRNA sequences define the ends of the island: the full length tRNA<sup>PheU</sup> (77 bp) is on the right of the figure; the left-hand partial tRNA (52 bp) and the central partial tRNA (16 bp) are thought to be att phage attachment sites. Full sequences of SPI-7 from other serovars are not yet available. PCR, mapping and partial sequence data indicate that *S. Dublin* and *S. Paratyphi C* carry very similar islands (>99 % nucleotide identity and absolute synteny), but both lack SopE phage and the region between the 16 bp and 77 bp tRNAs, giving islands of approximately 90 kb. Some strains of *S. Paratyphi C* contain a deletion within the *pil* locus [3-5].



The insertion of such a large element into a bacterial genome causes an imbalance in the replication geometry, meaning that the origin (*ori*) and terminus (*ter*) of replication are no longer positioned 180° apart, which may adversely affect replication [1]. Other *Salmonella* serovars demonstrate a more conserved gene order than *S. Typhi*, indicating that their genome structure is more stable [14]. It is proposed that this instability in the genome is a consequence of the insertion of islands such as SPI-7. Populations of *S. Typhi* cells have been observed to undergo spontaneous

genome rearrangements, and those which are closest to restoring *ori-ter* polarity have the most stable genomes and fastest generation times [14].

### Mobility of SPI-7

SPI-7 carries many predicted coding sequences (CDSs) with homology to genes involved in plasmid replication and transfer. There are also putative phage integrases (two in the *S. Typhi* island, and one in the *S. Dublin* and *S. Paratyphi C* islands) which may promote the excision and integration of the island. A genomic island from *Haemophilus influenzae*, *ICEHin1056*, has notable similarity to SPI-7 [15]. *ICEHin1056* has been found to encode a novel Type IV Secretion System, which allows conjugation of the island [16]. It is likely that the homologous region in SPI-7 has a similar role.

These lines of evidence imply that SPI-7 may be self-transmissible, although attempted transfer of the island to a donor strain by conjugation was unsuccessful [17]. Mobilisation of a non-self-transmissible plasmid, using the SPI-7 transfer region, has been demonstrated [17], indicating that *S. Typhi* can use the genes from SPI-7 in conjugation. The region from STY4554 to STY4586 is implicated in this activity. SPI-7 may have lost the ability to mobilise itself through a mutated gene at another locus.

### Vi Capsule Polysaccharide Operon

The Vi capsule polysaccharide was first identified as being associated with virulence in mice, and as an antigen distinct from that of O or H antigens [18,19]. Structurally, it is a polymer of N-acetylaminohexuronic acid [20].

Two genetic loci, *viaA*-locus and *viaB*-locus, are required for the production of Vi antigen in *S. Typhi* and *S. Paratyphi C* [21,22]. Whereas *viaB*-locus is uniquely present in Vi-positive strains, the *viaA*-locus operon is carried by many *Salmonella* serovars and *E. coli*, and has been shown to encompass the two-component regulatory system also involved in *E. coli* capsule synthesis, *rcsB-C* [23,24]. The *viaB*-locus operon has been localised to SPI-7 [25] and comprises a region responsible for the biosynthesis of the polysaccharide, and a region involved in its export [26,27]. Nucleotide sequencing of this locus determined that the biosynthesis operon consists of 5 CDSs, *tviA-E*, and the export is performed by the products of

*vexA-E* [28,29]. Much effort has gone into the elucidation of the function of the Vi antigen, although the picture remains unclear. Its roles in pathogenesis and immune response are discussed in two reviews [30,31].

Vi has become a focus for the production of vaccines. The efficacy and safety of these is reviewed [32-35], and development of new typhoid Vi vaccines is ongoing [36,37]. The future of Vi-based vaccines will depend on the maintenance of Vi within populations of *S. Typhi*; there are worrying reports that the *viaB*-locus can be inactivated in a number of ways. As mentioned above, spontaneous excision of SPI-7 has been observed, which would result in loss of the Vi phenotype [9,11,12]. Several other strains of *S. Typhi* and *S. Dublin* have been found to be Vi-negative, yet still maintain SPI-7. These are likely to carry mutations within the *viaA*-locus or *viaB*-locus [4,12]. Although most of these mutations have occurred after storage or passage of strains, a study of strains directly isolated from the blood of typhoid patients in Pakistan detected loss of SPI-7, or *viaB*-locus deletion [38]. If these strains are still exhibiting typhoid-like behaviour in the absence of Vi antigen, it implies that Vi is not essential to the development of typhoid fever, and that *S. Typhi* can potentially circumvent Vi-based vaccines.

### SopE Phage and Effector

The SopE phage carries the gene encoding SopE effector protein. SopE was first identified in *S. Dublin* as a 30 kDa secreted effector protein [39], translocated into target host cells through the Type III Secretion System encoded on SPI-1 [40]. Once inside the cell, SopE interacts with Rho-GTPase signalling molecules, promoting guanosine nucleotide exchange [41]. The activated signalling pathways lead to actin rearrangements within the cell, causing membrane ruffling and aiding bacterial invasion [41].

The phage itself is cryptic, similar to P2 phage [40], and inducible [42]. It is present within SPI-7 in strains of *S. Typhi*, but absent from the *S. Paratyphi C* and *S. Dublin* versions of SPI-7, and is found in other strains of *S. enterica* subsp. *enterica* in different genomic contexts. As SopE phage is not exclusively linked to SPI-7, the effects of SopE cannot be correlated to the systemic pathology of SPI-7-carrying strains, but may be

involved in cell invasion during gastrointestinal infection.

### Type IVB Pilus Production

SPI-7 carries an operon responsible for the production of type IVB pili (*pilL-pilV*; STY4539-STY4550) [43]. The pili have a diameter of 6nm, as observed from an overexpressing construct under electron microscopy [43], and are expressed primarily during stationary phase in liquid culture [44].

The pili were initially characterised as enhancing the adhesion to and/or invasion of human epithelial cells: in both *S. Typhi* and *S. Dublin*, knocking out the major structural prepilin subunit, PilS, was observed to reduce bacterial uptake by target cells [4,43,45]. This protein, PilS, also appears to mediate the interaction with the chloride channel CFTR (Cystic Fibrosis Transmembrane Conductance Regulator) [46], which is involved in the uptake of *S. Typhi* into epithelial cells [47]. More recent data shows that a *pilS* knockout mutant does not suffer from reduced adhesion to, or invasion of, epithelial cells [48], indicating that the interaction is complex and may vary depending on the specific strain or growth conditions. In addition, the type IVB pili are implicated in increasing inflammatory response in human monocytic cells [49].

A second role for the pili seems to be in bacterial self-association. The *pilV* gene product has been putatively identified as a pilus adhesin protein, as in the R64 thin pilus [50]. A shufflon has been observed in both systems, with the Rci recombinase performing a switch between possible alternative C termini of the PilV protein, of which there are two in SPI-7 [51]. In a *pilV* knockout, the *S. Typhi* cells are seen to self-associate, whereas constitutive expression of *pilV*, with either C terminus, reduces this effect [45]. It is thought that when the shufflon is switched rapidly, as may occur under certain growth conditions, full length PilV is not produced and the bacteria may self-associate [45]. The autoaggregation phenomenon is also observed in *S. Dublin* [4] but is absent from *S. Paratyphi C*, in which the Rci recombinase cannot operate due to mutation in its target sites, and the PilV is locked with an invariant C terminus [52]. In *S. Paratyphi C*, it is proposed that this mutation may be responsible for rendering the strains less likely to cause epidemics, as

compared to *S. Typhi* strains. It should be noted that some strains of *S. Paratyphi C* have deletions within the *pil* operon, from *pilS* (STY4547) to *rci* (STY4552) [5].

### Evolutionary Origins of SPI-7

As well as *Haemophilus influenzae*, discussed above, a number of organisms have been shown to harbour genomic islands which share genes and synteny with SPI-7: *Xanthomonas axonopodis*, *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, *Burkholderia xenovorans*, *Ralstonia metallidurans*, *Yersinia enterocolitica* and *Photobacterium luminescens* [5,15]. The level of amino acid identity is often very low, but core elements of the islands can be identified, including portions of the *rep*, *tra* and *int* regions (Figure). These islands carry varied cargoes, with the SPI-7 *viaB*-locus operon being replaced by genes encoding antibiotic resistance cassettes, toxins and metabolic proteins, reflecting the differences between these bacteria in habitat and selective pressure. This leads to the appealing idea of an ancient ancestral element, which has undergone extensive divergence, picking up alternative functional modules during its dispersal.

### Summary

SPI-7 represents an intriguing pathogenicity island, part of a family of islands found in a broad range of Gram negative bacteria. Its functions have yet to be fully determined, especially those of the associated virulence factors. No strict correlation between a specific factor and the ability to cause systemic disease has been found, confounded by the ability of *S. Paratyphi A* to cause enteric disease without SPI-7. Vi antigen holds potential for a typhoid vaccine, but the discovery of Vi-negative typhoid causing strains invites caution.

### Acknowledgements

This work was supported by the Wellcome Trust, through its core funding for the Wellcome Trust Sanger Institute. Thanks go to John Wain for critical review of the manuscript.

### References

1. Liu SL and Sanderson KE (1995) Rearrangements in the Genome of the Bacterium *Salmonella typhi*. *Proc Natl Acad Sci* 92(4):1018-1022.
2. Parkhill J *et al.* (2001) Complete Genome Sequence of a Multiple Drug Resistant *Salmonella enterica* serovar Typhi CT18. *Nature* 413(6858) 848-852.
3. Liu WQ *et al.* (2007) Diverse Genome Structures of *Salmonella paratyphi C*. *BMC Genomics* 8:290-299.
4. Morris C *et al.* (2003) *Salmonella enterica* serovar Dublin strains which are Vi Antigen-Positive use Type IVB Pili for Bacterial Self-Association and Human Intestinal Cell Entry. *Microb Pathogenesis* 35(6):279-284.
5. Pickard D *et al.* (2003) Composition, Acquisition, and Distribution of the Vi Exopolysaccharide-Encoding *Salmonella enterica* Pathogenicity Island SPI-7. *J Bacteriol* 185(17):5055-65.
6. Deng W *et al.* (2003) Comparative Genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18. *J Bacteriol* 185(7):2330-7.
7. McClelland M *et al.* (2004) Comparison of Genome Degradation in Paratyphi A and Typhi, Human-Restricted serovars of *Salmonella enterica* That Cause Typhoid. *Nat Genet* 36(12):1268-1274.
8. Hacker J *et al.* (1997) Pathogenicity Islands of Virulent Bacteria: Structure, Function and Impact on Microbial Evolution. *Mol Microbiol* 23(6):1089-1097.
9. Bueno SM *et al.* (2004) Precise Excision of the Large Pathogenicity Island, SPI7, in *Salmonella enterica* serovar Typhi. *J Bacteriol* 186(10):3202-13.
10. Hansen-Wester I and Hensel M (2002) Genome-Based Identification of Chromosomal Regions Specific for *Salmonella* spp. *Infect Immun* 70(5):2351-2360.
11. Wain J *et al.* (2005) Vi Antigen Expression in *Salmonella enterica* serovar Typhi Clinical Isolates from Pakistan. *J Clin Microbiol* 43(3):1158-65.
12. Nair S *et al.* (2004) *Salmonella enterica* serovar Typhi strains from which SPI7, a 134-Kilobase Island with Genes for Vi Exopolysaccharide and Other Functions, has been Deleted. *J Bacteriol* 186(10):3214-23.
13. Boyd EF *et al.* (2003) Differences in Gene Content among *Salmonella enterica* serovar Typhi Isolates. *J Clin Microbiol* 41(8):3823-8.
14. Liu GR *et al.* (2006) Genome Plasticity and *ori-ter* Rebalancing in *Salmonella typhi*. *Mol Biol Evol* 23(2):365-71.
15. Mohd-Zain Z *et al.* (2004) Transferable Antibiotic Resistance Elements in *Haemophilus influenzae* Share a Common Evolutionary Origin with a Diverse Family of Syntenic Genomic Islands. *J Bacteriol* 186(23):8114-8122.
16. Juhas M *et al.* (2007) Novel Type IV Secretion System Involved in Propagation of Genomic Islands. *J Bacteriol* 189(3):761-771.
17. Baker S *et al.* (2008) Mobilization of the *incQ* Plasmid R300B with a Chromosomal Conjugation System in *Salmonella enterica* serovar Typhi. *J Bacteriol* 190(11):4084-7.
18. Felix A and Pitt RM (1934) Virulence of *B. typhosus* and Resistance to O Antibody. *J Path Bact.* 38:409-420.
19. Felix A and Pitt R (1934) A New Antigen of *B. typhosus*. *Lancet* 227:186-191.
20. Clark WR, McLaughlin J, Webster ME (1958) An Aminohexuronic Acid as the Principal Hydrolytic Component of the Vi Antigen. *J Biol Chem* 230:81-89.
21. Johnson EM Krauskopf B, Baron LS (1965) Genetic Mapping of Vi and Somatic Antigenic Determinants in *Salmonella*. *J Bacteriol* 90(2):302-308.

22. Snellings NJ, Johnson EM, Baron LS (1977) Genetic Basis of Vi Antigen Expression in *Salmonella paratyphi* C. *J Bacteriol* 131(1):57-62.
23. Houg HS *et al.* (1992) Expression of Vi Antigen in *Escherichia coli* K-12: Characterization of ViaB from *Citrobacter freundii* and Identity of ViaA with RcsB. *J Bacteriol* 174(18):5910-5915.
24. Virlogeux I *et al.* (1996) Characterization of the *rcaA* and *rcaB* Genes from *Salmonella typhi*: *rcaB* through *tvIA* Is Involved in Regulation of Vi Antigen Synthesis. *J Bacteriol* 178(6):1691-1698.
25. Liu SL and Sanderson KE (1995) Genomic Cleavage Map of *Salmonella typhi* TY2. *J Bacteriol* 177(17):5099-5107.
26. Hashimoto Y (1991) *et al.* Molecular Cloning of the ViaB Region of *Salmonella typhi*. *FEMS Microbiol Lett* 69(1):53-56.
27. Kolyva S, Waxin H, Popoff MY (1992) The Vi Antigen of *Salmonella typhi*: Molecular Analysis of the *viaB* Locus. *J Gen Microbiol* 138(2):297-304.
28. Hashimoto Y *et al.* (1993) Complete Nucleotide Sequence and Molecular Characterization of ViaB Region Encoding Vi Antigen in *Salmonella typhi*. *J Bacteriol* 175(14):4456-4465.
29. Virlogeux I *et al.* (1995) Role of the *viaB* locus in synthesis, transport and expression of *Salmonella typhi* Vi antigen. *Microbiology-UK*. 141:3039-3047.
30. Raffatellu M *et al.* (2006) Capsule-Mediated Immune Evasion: a New Hypothesis Explaining Aspects of Typhoid Fever Pathogenesis. *Infect Immun* 74(1):19-27.
31. Robbins JD and Robbins JB (1984) Reexamination of the Protective Role of the Capsular Polysaccharide (Vi Antigen) of *Salmonella typhi*. *J Infect Dis* 150(3):436-449.
32. Hessel L *et al.* (1999) Experience with *Salmonella typhi* Vi Capsular Polysaccharide Vaccine. *Eur J Clin Microbiol Infect Dis* 18(9):609-620.
33. Ivanoff B and Levine MM (1997) Typhoid Fever: Continuing Challenges from a Resilient Bacterial Foe. *Bulletin De L'Institut Pasteur* 95(3):129-142.
34. Levine MM Tacket CO, Sztein MB (2001) Host-*Salmonella* Interaction: Human Trials. *Microb Infect* 3(14-15):1271-1279.
35. Parry CM (2004) Typhoid Fever. *Curr Infect Dis Rep*. 6(1):27-33.
36. Azze RFO *et al.* (2003) Immunogenicity of a New *Salmonella typhi* Vi Polysaccharide Vaccine-vax-TyVi (R) - in Cuban School Children and Teenagers. *Vaccine* 21(21-22):2758-2760.
37. Tacket CO *et al.* (2004) Immune Responses to an Oral Typhoid Vaccine Strain that is Modified to Constitutively Express Vi Capsular Polysaccharide. *J Infect Dis* 190(3):565-70.
38. Baker S *et al.* (2005) Detection of Vi-Negative *Salmonella enterica* serovar Typhi in the Peripheral Blood of Patients with Typhoid Fever in the Faisalabad Region of Pakistan. *J Clin Microbiol* 43(9):4418-25.
39. Wood MW *et al.* (1996) SopE, a Secreted Protein of *Salmonella dublin*, is Translocated into the Target Eukaryotic Cell via a Sip-Dependent Mechanism and Promotes Bacterial Entry. *Mol Microbiol* (22(2):327-338.
40. Hardt WD, Urlaub H, Galan JE (1998) A Substrate of the Centisome 63 Type III Protein Secretion System of *Salmonella typhimurium* is Encoded by a Cryptic Bacteriophage. *Proc Natl Acad Sci*. 95(5):2574-2579.
41. Hardt WD *et al.* (1998) *S. typhimurium* Encodes an Activator of Rho GTPases that Induces Membrane Ruffling and Nuclear Responses in Host Cells. *Cell* 93(5):815-826.
42. Mirolid S *et al.* (1999) Isolation of a Temperate Bacteriophage Encoding the Type III Effector Protein SopE from an Epidemic *Salmonella typhimurium* strain. *Proc Natl Acad Sci* 96(17):9845-9850.
43. Zhang XL *et al.* (2000) *Salmonella enterica* serovar Typhi uses Type IVB Pili to Enter Human Intestinal Epithelial Cells. *Infect Immun* 68(6):3067-73.
44. Lee FK, Morris C, Hackett J (2006) The *Salmonella enterica* serovar Typhi Vi Capsule and Self-Association Pili Share Controls on Expression. *FEMS Microbiol Lett* 261(1):41-46.
45. Morris C *et al.* (2003) The Shufflon of *Salmonella enterica* serovar Typhi Regulates Type IVB Pilus-Mediated Bacterial Self-Association. *Infect Immun* 71(3):1141-1146.
46. Tsui ISM *et al.* (2003) The Type IVB pili of *Salmonella enterica* serovar Typhi Bind to the Cystic Fibrosis Transmembrane Conductance Regulator. *Infect Immun* 71(10):6049-6050.
47. Pier GB *et al.* (1998) *Salmonella typhi* uses CFTR to Enter Intestinal Epithelial Cells. *Nature* 393(6680):79-82.
48. Bishop A *et al.* (2008) Interaction of *Salmonella enterica* serovar Typhi with Cultured Epithelial Cells: Roles of Surface Structures in Adhesion and Invasion. *Microbiology* 154(7):1914-1926.
49. Wang F *et al.* (2005) Type IVB Piliated *Salmonella typhi* Enhance IL-6 and NF-kappaB Production in Human Monocytic THP-1 Cells Through Activation of Protein Kinase C. *Immunobiology* 210(5):283-93.
50. Komano T *et al.* (1994) DNA Rearrangement of the Shufflon Determines Recipient Specificity in Liquid Mating of Inc11 Plasmid R64. *J Mol Biol* 243(1):6-9.
51. Zhang XL, Morris C, Hackett J (1997) Molecular Cloning, Nucleotide Sequence, and Function of a Site-Specific Recombinase Encoded in the Major 'Pathogenicity Island' of *Salmonella typhi*. *Gene* 202(1-2):139-46.
52. Tam CK, Hackett J, Morris C (2004) *Salmonella enterica* serovar Paratyphi C Carries an Inactive Shufflon. *Infect Immun* 72(1):22-8.
53. Carver TJ *et al.* (2005) ACT: the Artemis Comparison Tool. *Bioinformatics*. 21:3422-3423.

**Corresponding Author:** Helena M. B. Seth-Smith, Pathogen Sequencing Unit, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom. Tel: (+44) 1223 834244, Fax: (+44) 1223 494919. E-mail: hss@sanger.ac.uk

**Conflict of interest:** No conflict of interest is declared.