Genomic Comparison of the Closely Related *Salmonella enterica* Serovars Enteritidis and Dublin

Laura Betancora, Lucía Yima, Arací Martíneza,b, Maria Fookesc, Sebastian Sasiasa, Felipe Schelottob, Nicholas Thomasonc, Duncan Maskeld and José A. Chabalgoitya,*

aDepartment of Desarrollo Biotecnológico and bDepartment of Bacteriología y Virología, Instituto de Higiene, Facultad de Medicina, Universidad de la República, Av. A. Navarro 3051, CP 11600, Montevideo, Uruguay; cThe Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; dDepartment of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge CB3 0ES, UK

Abstract: The Enteritidis and Dublin serovars of *Salmonella enterica* are closely related, yet they differ significantly in pathogenicity and epidemiology. *S. Enteritidis* is a broad host range serovar that commonly causes gastroenteritis and infrequently causes invasive disease in humans. *S. Dublin* mainly colonizes cattle but upon infecting humans often results in invasive disease. To gain a broader view of the extent of these differences we conducted microarray-based comparative genomics between several field isolates from each serovar. Genome degradation has been correlated with host adaptation in *Salmonella*, thus we also compared at whole genome scale the available genomic sequences of them to evaluate pseudogene composition within each serovar.

Microarray analysis revealed 3771 CDS shared by both serovars while 33 were only present in Enteritidis and 87 were exclusive to Dublin. Pseudogene evaluation showed 177 inactive CDS in *S. Dublin* which correspond to active genes in *S. Enteritidis*, nine of which are also inactive in the host adapted *S. Gallinarum* and *S. Choleraesuis* serovars. Sequencing of these 9 CDS in several *S. Dublin* clinical isolates revealed that they are pseudogenes in all of them, indicating that this feature is not peculiar to the sequenced strain. Among these CDS, *shdA* (Peyer’s patch colonization factor) and *mglA* (galactoside transport ATP binding protein), appear also to be inactive in the human adapted *S. Typhi* and *S. Paratyphi* A, suggesting that functionality of these genes may be relevant for the capacity of certain *Salmonella* serovars to infect a broad range of hosts.

Keywords: Comparative genomics, Host specificity, Pseudogenes, *Salmonella*, *S. Dublin*, *S. Enteritidis*.

1. **INTRODUCTION**

Infection with non-typhoidal *Salmonella enterica* is a major cause of food-borne disease in humans worldwide [1-3]. Animals and their products are regarded as the main sources of this pathogen, although it may also be present in other potential sources, such as fresh vegetables [4-6]. From over 2500 different serovars of *Salmonella enterica* (defined by their surface antigenic properties, both somatic O antigen and flagellar H antigens) about 50 are significant pathogens of animals and humans. Acute infections in humans can develop in one of four ways: enteric fever, gastroenteritis, bacteremia, or extraintestinal focal infection [7]. As with other infectious diseases, the course and outcome of the infection depend on a variety of factors, including the inoculating dose, the immune status of the host, and the genetic background of both the host and the infecting organism.

Although *S. enterica* serovars are genetically very similar, they differ significantly in host range and disease spectrum. *S. enterica* serovars may be classified as ubiquitous, host-restricted or host-specific. Ubiquitous serovars, which include Typhimurium and Enteritidis, most commonly produce self-limiting gastrointestinal infections in a wide range of hosts. Host-specific serovars, such as Typhi in humans or Gallinarum in fowl, cause severe systemic diseases in their specific hosts. A few *Salmonella* serovars, such as Choleraesuis and Dublin, have a narrow host range and are classified as host-restricted [8].

Host-restricted and host-specific serovars are generally more prone to cause invasive disease than ubiquitous serovars [9, 10]. Globally, human extra-intestinal salmonellosis is generally associated with those serovars that are also associated with gastroenteritis, as is the case with *S. Enteritidis* and *S. Typhimurium*. However, certain serovars are more prone to cause invasive infections than others, as is clear when the percentage of isolates from bacteremia related to total cases (invasive index) is calculated [7, 11]. For *S. Typhimurium* and *S. Enteritidis*, the invasive index ranges from 1 to 7% [11, 12], while for *S. Dublin* different reports indicate that the invasive index ranges from 50% to 70% [7, 11, 13-15]. Loss of gene function through pseudogene accumulation has been indicated as a hallmark of host-
specific pathogenic bacteria as compared to their host-
generalist relatives [16-22].

The Enteritidis (O: 1, 9, 12; gm: -) and Dublin (O:1, 9, 12; gp: -) serovars share antigenic properties and are phylogenetically closely related, yet they seem to differ significantly in pathogenic potential [23, 24]. S. Enteritidis commonly causes gastroenteritis but rarely causes invasive disease in humans. S. Dublin usually infects cattle causing abortion and systemic infection, but occasionally can be found infecting other hosts such as pigs and humans. On the rare occasions when it infects humans it often results in bacteraemia with severe disease and high mortality [25-27]. Characterization of the mechanisms underlying these differences is central to a more general understanding of the invasiveness of salmonellae. To date only one complete genome of a S. Enteritidis strain (P125109, hereafter referred as PT4) and two S. Dublin isolates (CT_02021853 and 3246) have been sequenced and annotated and are publicly available [28], [http://www.ncbi.nlm.nih.gov/genomeprj/19467] [29].

To gain new insights into genetic differences that could help to understand the basis of such marked different pathogenic behaviors, here we describe a comparative study between S. Enteritidis and S. Dublin. We conducted microarray-based comparative genomics between four S. Dublin clinical isolates and the core genome resulting of the comparative genome analysis of 29 S. Enteritidis isolates previously reported by us [30]. Further the pseudogene content of each serovar was also evaluated using the available genome sequences.

2. MATERIALS AND METHODS

2.1. Bacterial Strains

Twenty-nine S. enterica serovar Enteritidis isolates from diverse origins in Uruguay were previously characterized by microarray and phenotypic assays [30, 31]. Seven S. enterica serovar Dublin isolates from human infections in Uruguay were used in this study (Table 1).

Isolates were maintained frozen at -80°C in LB containing 25% glycerol. Bacteria were cultured in LB broth, or on LB containing 1.6% agar, or Tryptic Soy Agar. All isolates were identified as Salmonella enterica using standard biochemical tests and microbiological methods.

Serovar was determined by the slide agglutination test for O antigen and the tube agglutination test for H antigen, using commercially available anti-O and anti-H antisera (Difco, France). Differentiation between S. Enteritidis and S. Dublin was confirmed by PCR for the detection of genetic regions specific for Enteritidis [32] and by sequencing the fliC gene, which differs between these serovars.

2.2. Comparative Genomic Hybridization Analysis (CGH)

Four S. Dublin strains were analyzed by CGH using the Salmonella generation IV microarray [30, 33, 34] with PT4 DNA [28] as reference. The array is non-redundant and contains coding sequences from the following eight genomes: S. enterica serovar Typhi (S. Typhi) CT18, S. Typhi Ty2, S. Typhimurium LT2 (ATCC 700220), S. Typhimurium DT104 (NCTC 13348), S. Typhimurium SL1344 (NCTC 13347), S. Enteritidis PT4 P125109 (NCTC 13349), S. Gallinarum 287/91 (NCTC 13346), and S. bongori 12419 (ATCC 43975). Total DNA (including plasmid DNA) was extracted from each strain using a Genome DNA extraction kit (Promega) and quantified by agarose gel electrophoresis. Labeled DNA from S. Enteritidis PT4 (control sample) and one of the query Salmonella strains (experimental sample) were mixed in equal volumes and concentrations and hybridized to the microarray slides as previously described [30]. Data were normalized to the median value, and the total list of 6,871 genes was filtered by removing those spots with a high background and those without data in at least one of the replicates (three slides per strain, duplicate features per slide). After filtering, a list of 5,695 genes was obtained that corresponded to genes that presented a valid signal in at least one of the strains analyzed. Data analysis was performed on Excel files, following criteria previously described [30].

Genes assigned as absent/divergent in all S. Dublin isolates were compared to the core genome of S. Enteritidis as defined in our previous study [30]. Genes detected as present in all S. Dublin isolates but absent in S. Enteritidis PT4 were compared with the S. Enteritidis dispensable genome as well as with the fully sequenced Salmonella isolates available in the NCBI database. Genes encoded in plasmids were not considered in this analysis.

Table 1. Description of the S. Dublin Isolates Analyzed in This Work

| Strain Designation | Year of Isolation | Origin | CGH | 9 CDS Sequence |
|--------------------|-------------------|--------|-----|---------------|
| SDU1               | 1995              | blood  | +   | +             |
| SDU2               | 2004              | blood  | +   | +             |
| SDU3               | 2006              | blood  | +   | +             |
| SDU4               | 2008              | blood  | -   | +             |
| SDU5               | 2000              | feces  | +   | +             |
| SDU6               | 2005              | feces  | -   | +             |
| SDU7               | 2008              | feces  | -   | +             |

+: tested. -: non-tested. *Correspond to human samples. *Comparative genomics hybridization. *Nucleotide sequence of CDS as described in text and Table 2.
2.3. Web Based Comparative Genomics

The sequences and annotations of the *Salmonella* genomes analyzed here were obtained from the data available at NCBI [http://www.ncbi.nlm.nih.gov/]. Nucleotide sequences were analyzed using the sequence visualization and annotation tool Artemis version 10 [35]. The search for homologous genes and regions was performed using Blast-n and Blast-p online at the NCBI website.

2.4. Pseudogene Screening in *S. Dublin* Isolates

The sequences of nine CDS detected as pseudogenes in the *S. Dublin*, *S. Gallinarum* and *S. Choleraesuis* sequenced strains were evaluated in all 7 *S. Dublin* isolates included in this work. Genomic DNA was extracted from the bacterial strains using DNeasy blood and tissue kit (Qiagen). Specific primers for amplification and sequencing were designed based on the sequences of the corresponding regions in the genomes of *S. Enteritidis* PT4 and *S. Dublin* CT_0201853 (Table 2). PCRs were conducted using a 10:1 mix, in terms of units, of *Taq* Polymerase and *Pfu* Polymerase (Fermentas) and the PCR products were sequenced. Sequences were analyzed and aligned using BioEdit Sequence Alignment editor version 7.0.9.0, 2007.

Table 2. Description of the Primers used for Amplifying and Sequencing the 9 CDS Described as Pseudogenes in *S. Gallinarum*, *S. Dublin* and *S. Choleraesuis* Fully Sequenced Strains

| Gene in *S. Enteritidis* | Primer Sequence (5´-3´) | Reverse/Forward |
|--------------------------|--------------------------|-----------------|
| SEN0042                  | TATTCAAAACTTGCTTAGAAAGTAGAG | Forward |
|                          | CGGGTCTTGGTGCATAAATGG | Reverse |
|                          | GGAAGTAAATGGTGGTCCGGTG | Reverse2 |
| SEN0784                  | GTGGTAAACATATATGTAATGTATTTTC | Forward |
|                          | AAATGTTGATCCAGGTGGTCT | Reverse |
| SEN2182                  | AGACCGGATAAAGCTATTTGTCTTTGCC | Forward |
|                          | ATTCGCCCATTTTCAGCAGCTG | Reverse |
|                          | GTGATTGTCAGCCGCGGCTCTCTC | Reverse2 |
| SEN2493                  | TCCAGTTTGCTTCGTGAACG | Forward |
|                          | CACTGGCGATGTCAGCCTT | Forward2 |
|                          | CAATTTGGCCGTAATGACGGT | Forward3 |
|                          | ATCAACCGGTGTTGTCACTG | Reverse |
|                          | TACCGTCCGAGTCGCGGTGG | Reverse2 |
| SEN2783                  | GTTAGGTATATCAACAAAAAGACCA | Forward |
|                          | TCCAGAGCAGCCATCCAGGA | Forward2 |
|                          | TGTGCAAGCCCAAGCCTG | Forward3 |
|                          | ACGGAGGGAGCCCAGG | Reverse |
|                          | CAAACCTCTTGCCGTTGTAATCACC | Reverse2 |
| SEN2806                  | GTGCTGTAGGCGATATTTAAG | Forward |
|                          | CTTCGCCGAGGCGGCTAT | Forward2 |
|                          | AACCTGCAATTCAGCTACTACAG | Reverse |
| SEN3461                  | TTTGACAGGCTCGCCGACAT | Forward |
|                          | GAATGCCTGTCGGTGGATT | Forward2 |
|                          | CGTGCCGGGAATCAAACAG | Forward3 |
|                          | AGCACCAGCCGGCCCACA | Reverse |
|                          | GCCGCGAAAACCGTAGTTCG | Reverse2 |
| SEN3672                  | GGGCTGTCAGCCTGTAAC | Forward |
|                          | CTCTTTTTGCCTTGATCC | Forward2 |
|                          | TATGACGGTTGTGATGCAATGG | Reverse |
| SEN4290                  | AACGCTTGAGGATTTAATAGAA | Forward |
|                          | CTGATTTCAGTACCGTCAG | Reverse |
3. RESULTS

3.1. Microarray-based Comparative Genomics of S. Enteritidis and S. Dublin Isolates

The genetic content of the 4 S. Dublin isolates was evaluated by microarray and a core genome (i.e. genes present in all strains) was defined. To explore the genetic determinants underlying the phenotypic differences between S. Dublin and S. Enteritidis, we compared the core genome of S. Dublin with the previously defined core genome of S. Enteritidis [30]. We found 3771 genes shared by both serovars, whereas 33 genes were only present in S. Enteritidis strains (Table 3) and 87 genes were only present in S. Dublin isolates (Table 4). The regions of difference found by CGH analysis are similar to the regions of difference obtained from comparison of the genomes of the two sequenced strains PT4 and CT_02021853 (results not shown). From these 120 (33 + 87) genes which are exclusive of one serovar or the other, 53 are bacteriophage-encoded.

As shown in Table 3 four DNA regions and seven single genes were present only in S. Enteritidis. Region En1 (SEN0083-SEN0085) encodes two putative secreted proteins and one sulphatase. BLAST analysis revealed that this region has homologues in several fully sequenced serovars of Salmonella, including S. Gallinarum, S. Typhi, S. Paratyphi A, S. Paratyphi B, S. Choleraesuis, S. Typhimurium, S. Agona, S. Newport and S. Heidelberg. Region En2 (SEN1379-1395), corresponds to phage SE14 [28], that includes genes encoding for DNA nucleases and membrane proteins, and was previously postulated to be a region of difference between S. Enteritidis and all other Salmonella serovars [30, 36]. Region En3 (SEN1432-1435) corresponds to a genomic island previously described as ROD13 [28] that encodes for idonate dehydrogenase, gluconate dehydrogenase, proteins involved in sugar transport, and proteins similar to those required for hexonate uptake. This genomic island is present in the S. Gallinarum genome sequence, but is absent from all other salmonellae sequenced to date. Region En4 (SEN1500-1506) corresponds to part of another genomic region, named ROD14 [28], and encodes for a putative transcriptional regulator akin to the LacI family, and other regulatory proteins probably involved in drug efflux. This region is present in the genome sequences from various S. Typhimurium strains, but is degraded in the S. Gallinarum and PT4 genome sequences.

Six regions and six isolated genes are present only in S. Dublin (Table 4). Region Du1 comprises thirteen genes previously annotated within the genome of S. Gallinarum (SG1032-1044) which include proteins that are members of the Rhs family, Clp proteases and exported proteins. Region Du2 (SG1182-1195 and SG1211-1219) corresponds to part of the Gifsy-2-like prophage remnant present in the genome of S. Gallinarum [28]. Region Du3 corresponds to genes found in SPI-6 from S. Typhi CT18.

Regions Du4, Du5 and Du6, correspond to prophages found in the genome sequence of S. Typhi CT18 [16]. Single genes present only in S. Dublin strains include a membrane transport protein (SG3368), a putative glycolate oxidase (STY1444) and several phase-related proteins.

Microarray methodology allowed us to detect only presence or absence/divergence of genes, but not small

---

Table 3. Regions (Reg) and Single Genes (Sing) that form the S. Enteritidis Core Genome but Appear as Absent/Divergent in S. Dublin Strains

| Region | Gene Range | Homologous | Function/Gene Prediction |
|--------|------------|------------|-------------------------|
| Reg En1 | SEN0083-0085 | CT18, TY2, LT2, DT104, SL1344, SBG, SPA, SGAL | probable secreted proteins, sulfatase |
| Reg En2 | SEN1379-1395 (1387 present) | STY (SOME) | part of PHAGE SE14, ligA, B, C, D, F, ydaD |
| Reg En3 | SEN1432-1435 | SGAL | ROD13 genomic island, idonate and gluconate dehydrogenase, sugar transport |
| Reg En4 | SEN1500-SEN1506 | LT2, SL1344, (CT18 and SBG some) | part of ROD14 genomic island |
| Sing En1 | SEN0196 | SBG | fhuA, ferrichrome iron receptor |
| Sing En2 | SEN0281 | NO | safA, fimbrial subunit |
| Sing En3 | SEN0356 | SGAL | putative autotransporter |
| Sing En4 | SEN1515 | CT18, TY2, LT2, DT104, SL1344, SBG, SPA, SGAL | Ni/Fe-hydrogenase 1 b-type cytochrome subunit HyaC2 |
| Sing En5 | SEN1539 | CT18, TY2, LT2, DT104, SL1344, SBG, SPA, SGAL | dcp, dipeptidil carboxipeptidaseII |
| Sing En6 | SEN2167 | CT18, TY2, LT2, DT104, SL1344, SBG, SPA, SGAL | conserved hypothetical protein |
| Sing En7 | SEN2420 | SGAL | putative exported protein |

CT18: S. Typhi CT18, TY2: S. Typhi Ty2, LT2: S. Typhimurium LT2, DT104: S. Typhimurium DT104, SL1344: S. Typhimurium SL1344, SBG: S. bongori, SPA: S. Paratyphi A, SGAL: S. Gallinarum.
Comparative Genomics among S. Enteritidis and S. Dublin

The Open Microbiology Journal, 2012, Volume 6

variations in gene sequences. Considering that pseudogene accumulation has been postulated to be involved in host restriction and adaptation, we decided to compare the pseudogene content among the available genomic sequences of both serovars and then evaluate if the Uruguayan S. Dublin clinical isolates harbour a particular set of these pseudogenes.

3.2. Pseudogene Analysis

Analysis of the genomes available in the NCBI database for S. Dublin CT_02021853 and S. Enteritidis PT4 strains, show that they have 289 CDS and 111 CDS annotated as pseudogenes respectively. From the 289 S. Dublin pseudogenes, 7 have no homologues in the S. Enteritidis sequence, and 32 correspond to intergenic regions. Among the others, 38 are homologous with 29 pseudogenes in S. Enteritidis, whereas the other 212 pseudogenes in S. Dublin correspond to 177 active genes in S. Enteritidis. Conversely, there are 83 S. Enteritidis pseudogenes that appear to be functional in S. Dublin CT_02021853. We analyzed the pseudogenes specific of each serovar, and grouped them in different classes according with their homology with functional CDS (Table 5).

S. Enteritidis, S. Dublin and S. Gallinarum form a related cluster of serovars but with marked differences in host-specificity, thus we also included S. Gallinarum in the pseudogene analysis. There is a single annotated genome sequence for this serovar that contains 309 pseudogenes [28] and among them only 21 are also annotated as pseudogenes in S. Dublin but not in S. Enteritidis (Table 6). This group of CDS includes nine that are also inactive (7) or completely absent (2) in the other host-restricted serovar S. Choleraesuis [37] and are described in Table 6.

Table 5. Distribution of the S. Enteritidis or S. Dublin Specific Pseudogenes among Different Functional Classes

| PseudogenesSEN (%) | PseudogenesSDU (%) |
|-------------------|--------------------|
| Surface           | 20.48              |
| methabolism       | 10.84              |
| regulatory        | 1.20               |
| transposase       | 15.66              |
| hypothetical      | 14.46              |
| Virulence         | 3.61               |
| ribosomal         | 0.00               |
| Phage             | 26.51              |
| Other             | 7.23               |

Table 4. Regions (Reg) and Single Genes (Sing) that are Present in all S. Dublin Strains but Absent in the S. Enteritidis Sequenced and Analyzed Isolates

| Gene Range   | Homologous | Gene Description                                      |
|--------------|------------|-------------------------------------------------------|
| Reg Du1      | SG1032-1044| NO                                                   |
| Reg Du2a     | SG1182-1195| SOME SDT, SOME STY                                   |
| Reg Du2b     | SG1211-1219| STM, SDT, SL                                         |
| Reg Du3a     | STY0289-0294| STM, SDT, SL, SPA, TY2, SOME GAL                     |
| Reg Du3b     | STY0302-0310| STM, SDT, SL, SPA, TY2                               |
| Reg Du3c     | STY0320-0323| STM, SDT, SL, SPA, TY2                               |
| Reg Du4      | STY1020-1036| TY2, SOME SDT, SDT, SL                               |
| Reg Du5      | STY2043-2045| SOME SDT                                             |
| Reg Du6      | STY3662-3671| TY2, SOME STM                                        |
| Sing Du1     | SG1227     | STM, SDT, SL                                         |
| Sing Du2     | SG3368     | STY, STM, SDT, SL, SBG, SPA                           |
| Sing Du3     | STY0602    | SDT, SBG, SPA                                        |
| Sing Du4     | STY1444    | TY2, STM, SDT, SL, SBG, SPA                           |
| Sing Du5     | STY2690    | TY2, STM, SDT,SL                                     |
| Sing Du6     | STY3029    | NO                                                   |

Overall, the presence of these nine pseudogenes could be regarded as potential distinguishing markers of host-
restricted serovars, thus we decided to evaluate their sequences in all S. Dublin Uruguayan isolates obtained from human infections (4 strains analyzed by CGH as described above plus 3 other isolates, Table 1). We found that all 7 isolates have these 9 CDS inactivated as pseudogenes, either by the same point mutations that are present in the fully sequenced S. Dublin CT_02021853 strain (7 of the 9 CDS) or by a different deletion as is the case of the CDS homologous to SEN2493 and SEN4290. Recently the genome sequence of another S. Dublin strain (S. Dublin 3246), was publicly released (GenBank: CM001151) [29]. We found that all 9 CDS are also pseudogenes in this strain. Further, in all but one of them the inactivation is due to the same changes than in S. Dublin CT_02021853. Interestingly, the exception is the CDS corresponding to SEN4290, which possess the same deletion than the Uruguayan strains analyzed here.

**DISCUSSION**

S. Enteritidis and S. Dublin are two closely related serovar with marked differences in pathogenic traits and epidemiological behavior, thus it is reasonable to assume that genomic comparison between them could shed some light on the molecular basis of these differences. A single previous report described a microarray-based genome comparison [38], and here we conducted a similar analysis using a different set of field isolates and microarray chip. Further, we now report a comparison of the full genome sequences of S. Enteritidis and S. Dublin particularly looking at differences in pseudogene composition between them.

Our comparative genome hybridization study predicted 33 genes specific to S. Enteritidis and 87 specific to S. Dublin. The analysis revealed four genetic regions and seven single genes that seem to be exclusive of S. Enteritidis core genome, as well as six regions and six single genes specific for S. Dublin. These results corroborate and extend the previous report where 3 S. Dublin and 24 S. Enteritidis strains where compared [38]. This report described the same four regions specific for Enteritidis but only one of the six S. Dublin regions found by us. This particular region, that we denominated Du3, corresponds to regions B24, B25_a and B25_b as of the earlier report. Region Du3 corresponds to genes found in SPI-6 from S. Typhi CT18. This region encodes a CtpB heat-shock protease-like protein, as well as

| Gene     | Choleraesuis Pseu/Absent* | Gene Description                           |
|----------|----------------------------|--------------------------------------------|
| SEN0042  | YES                        | putative transport protein                 |
| SEN0325  | NO                         | possible transmembrane regulator           |
| SEN0621  | NO                         | putative sigma54 dependent transcriptional regulator |
| SEN0784  | YES                        | hypothetical protein                       |
| SEN1194  | NO                         | putative membrane transport protein        |
| SEN1331  | NO                         | conserved hypothetical protein             |
| SEN1335  | NO                         | putative membrane protein                  |
| SEN1524  | NO                         | putative membrane protein                  |
| SEN2173  | NO                         | putative transcriptional regulator         |
| SEN2182  | YES                        | mgLa, galactoside transport ATP binding protein |
| SEN2493  | YES                        | shdA, Peyer’s patch colonization and shedding factor |
| SEN2611  | NO                         | putative type I secretion protein, SPI9 ATP-binding protein |
| SEN2783  | YES                        | conserved hypothetical protein             |
| SEN2806  | YES                        | ygeY probable glucarate dehydratase        |
| SEN3461  | YES                        | lpfC, outer membrane usher protein         |
| SEN3537  | NO                         | rfaZ (waaZ) LPS core biosynthesis protein  |
| SEN3571  | NO                         | yicJ sodium galactoside family symporter   |
| SEN3672  | YES                        | probable PTS system permease               |
| SEN3954  | NO                         | rfi, putative endonuclease V               |
| SEN4259  | NO                         | hypothetical protein                       |
| SEN4290  | YES                        | Type I restriction-modification system methyltransferase |

* YES indicates that the corresponding gene is a pseudogene or is absent in the genome of S. Choleraesuis SC-B67. NO indicates that corresponds to an active gene. a,b indicates that corresponds to a pseudogene in the sequences of S. Typhi CT18 and Ty2 as well as in S. Paratyphi A ATCC 9150 and S. Paratyphi A AKU_12601, as analyzed by Holt et al [22].
Comparative Genomics among S. Enteritidis and S. Dublin

Among the other regions specific for S. Dublin described here, Region Du1 was recently proposed to be a pathogenicity island (SPI-19) identified in S. Gallinarum, S. Dublin, S. Weltevreden and S. Agona that encodes a type-6 secretion system (T6SS) [39]. In S. Enteritidis, an internal deletion has eliminated most of the island. Region Du2, includes various bacteriophage regulatory proteins, recombinases, transposases, and structural proteins. It also includes one gene (SG1186) previously annotated as encoding a putative phage-encoded cell division inhibitor protein belonging to the kil super-family and associated with the capacity to inhibit the essential ftsZ cell-division gene [40]. ftsZ expression is altered during the intracellular phase of infection with S. enterica, a process that is independent of sulA, a known inhibitor of ftsZ [41]. Genes encoding proteins belonging to the same super-family are also present in several S. Typhi genome sequences, as well as in other enterobacteria (e.g. different STEC strains, Shigella flexneri, Shigella dysenteriae and others) as revealed by Blast-p analysis, suggesting a possible role for these proteins in pathogenesis. Regions Du1 and Du2 were not represented in the microarray used by Porwollik and collaborators [38], thus we cannot exclude that these regions were also present in the strains studied there, but simply not found because of the particular microarray used. Regions Du4, Du5 and Du6, correspond to prophages found in the genome sequence of S. Typhi CT18 [16]. Region Du4 comprises 17 genes from a lambdoid bacteriophage that include several CDS encoding for DNA binding proteins. Region Du5 includes 3 genes that are part of a degenerate bacteriophage; one of these (STY2044) encodes a putative endolysin similar to several lysozymes from E. coli and Shigella strains. Region Du6 spans 10 genes including a DNA adenine methylase (STY3667), regulatory proteins and endonucleases. These 3 regions of difference were not found in the earlier report, despite the CDS been present in the microarray. Instead, that work reported differences in other prophage-derived genes. Thus, it could that the genomes of the particular set of strains used in both studies possess different prophage composition. The analysis of Du4-Du5-Du6 in both S. Dublin sequenced isolates, revealed that regions Du5 and Du6 are very conserved in both strains whereas region Du4 is almost complete in CT_02021853 but incomplete and less conserved in strain 3246, supporting the hypothesis of different content in phage genes among S. Dublin isolates.

Among the seven single genes that are for the first time described here as absent in S. Dublin strains, safA (SEN0281) and dcp (SEN1539) are of special interest. safA is the first gene of the saf fimbrial operon and encodes a lipoprotein. The operon forms part of the degraded pathogenicity island SPI-6 in the S. Enteritidis chromosome. This operon is not annotated in the S. Dublin genome sequences available. However, Blast analysis revealed that this is a region highly conserved at a nucleotide level between PT4 and both S. Dublin sequenced isolates. There are several stop codons in the S. Dublin sequence homologous to safA, suggesting that this gene is in process of degradation. The fact that we cannot detect safA by CGH in the S. Dublin Uruguayan isolates may be related with this. The dcp gene encodes for dipeptidyl-carboxypeptidase II, which is highly conserved among the Enterobacteriaceae. This gene has been described previously as a frequent site for SNPs in S. Enteritidis [42], and it is absent from the CT_02021853 sequence.

Overall, the CGH analyses did not detect clear differences in genes that have been previously reported as required for virulence to explain the differences in pathogenicity of both serovars. However, the presence/absence of a gene, as detected by this methodology, does not inform about its expression, thus these results should be interpreted with caution.

The high number of pseudogenes detected in CT_02021853 suggests that this mechanism might be relevant in the process of host adaptation of this serovar, as well as in the different epidemiological and pathogenic behavior of S. Dublin when compared with S. Enteritidis. As we describe in Table 5, we observed a differential distribution of functionality amongst the CDS inactive in S. Enteritidis and S. Dublin. More than 40% of the pseudogenes specific for S. Enteritidis correspond to CDS related to phages or transposases but only 12% with those involved in metabolism and regulatory proteins. Conversely, among the pseudogenes specific for S. Dublin 33% correspond to CDS encoding proteins involved in central metabolism or regulatory proteins and 37% to CDS related to surface structures but only 3% to phages and transposases. These observations may be relevant to understand the host restriction of S. Dublin.

We found 21 CDS that appear to be active genes in the broad host-range S. Enteritidis but pseudogenes in the host-restricted S. Dublin and in the host-specific S. Gallinarum. From this set of CDS, 9 are pseudogenes as well in the other host-restricted serovar S. Cholerasuis suggesting that their inactivation could be relevant as genetic determinants of host adaptation. These nine CDS correspond to two hypothetical proteins (SEN0784 and SEN2783), one putative transport protein (SEN0042), the gene encoding the outer membrane usher protein LpIC (SEN3461), one probable phospho-transferase system permease (SEN3672), one gene encoding a putative Type I restriction modification system protein (SEN4290), and the gene encoding a probable glucarate dehydratase 2 (SEN2806 or vgcY). The other two genes that complete this list are mglA (SEN2182) and shdA (SEN2493), which are pseudogenes in S. Typhi CT18 and Ty2 as well as in S. Paratyphi A ATCC 9150 and S. Paratyphi A AKU_12601 [22]. ShdA is involved in colonization of Peyer’s patches by S. Typhimurium and in shedding of the bacteria after infection [43-45]. MglA is a galactoside transport ATP binding protein. The roles of these genes in the broad host-range of S. Enteritidis remain to be established.

All these nine CDS are pseudogenes in the seven S. Dublin clinical isolates evaluated in this work, as well as in the other fully sequenced isolate S. Dublin 3246, suggesting that the loss of their functionality is not a consequence of random mutation. Two of these 9 pseudogenes in the Uruguayan isolates have lost their functionality by mutations
that are different from those seen in the sequenced strain CT_02021853 suggesting that this loose of functionality involves a process of convergent evolution.

In conclusion, our results show several genetic differences that may help to explain why such close related organisms can nevertheless behave with such marked differences. Comparison of larger numbers of field strains at full genome scale is becoming increasingly feasible, and may provide new insights into the genetic basis of host adaptation.

CONFLICT OF INTEREST

None declared.

ACKNOWLEDGMENTS

This work was jointly supported by a project grant from the Wellcome Trust (078168/Z/05/Z) and by the Central Research Committee (CSIC) of the Universidad de la República Uruguay. We like to thanks Gordon Dougan and Derek Pickard for their helpful advices.

REFERENCES

[1] de Jong B, Ekdahl K. The comparative burden of salmonellosis in the European Union member states, associated and candidate countries. BMC Public Health 2006; 6: 4.
[2] Galanis E, Lo Fo Wong DM, Patrick ME, et al. Web-based surveillance and global Salmonella distribution, 2000-2002. Emerg Infect Dis 2006; 12: 381-8.
[3] Voetsch AC, Van Gilder TJ, Angulo FJ, et al. FoodNet estimate of the burden of illness caused by nontyphoidal Salmonella infections in the United States. Clin Infect Dis 2004; 38(Suppl 3): S127-34.
[4] Perales I, Audicana A. Salmonella Enteritidis and eggs. Lancet 1988; 2: 1133.
[5] Wells J, Butterfield J. Incidence of Salmonella on fresh fruits and vegetables affected by fungal rots or physical injury. Plant Dis 1999; 83: 722-6.
[6] Hald T, Vose D, Wegener HC, Koupeev T. A Bayesian approach to hybridization of sequences specific for Salmonella enterica subspecies I genovars by use of microarrays. J Bacteriol 2004; 186: 5883-98.
[7] Blaser MJ, Feldman RA. From the centers for disease control. Salmonella enteritaemia: reports to the Centers for Disease Control, 1968-1979. J Infect Dis 1981; 143: 743-6.
[8] Rutherford K, Parkhill J, Crook J, et al. Artemis: sequence and genotype visualization. Nat Biotechnol 2004; 22: 503-510.
[9] Anjum MF, Maroney C, F룸 kes M, et al. Identification of core and variable components of the Salmonella enterica species. Genome Res 2001; 11: 108-17.
Comparative Genomics among S. Enteritidis and S. Dublin

39 Blondel CJ, Jimenez JC, Contreras I, Santiviago CA. Comparative genomic analysis uncovers 3 novel loci encoding type six secretion systems differentially distributed in Salmonella serotypes. BMC Genomics 2009; 10: 354.

40 Conter A, Bouche JP, Dassain M. Identification of a new inhibitor of essential division gene ftsZ as the kil gene of defective prophage Rac. J Bacteriol 1996; 178: 5100-4.

41 Henry T, Garcia-Del Portillo F, Gorvel JP. Identification of Salmonella functions critical for bacterial cell division within eukaryotic cells. Mol Microbiol 2005; 56: 252-67.

42 Guard J. Evolutionary trends in two strains of Salmonella enterica subsp. Enteritidis PT13a that vary in virulence potential. Journal [serial on the Internet]. 2010 Date: Available from: http://www.ncbi.nlm.nih.gov/genomes/static/Salmonella_SNPS.html.

43 Kingsley RA, van Amsterdam K, Kramer N, Baumler AJ. The shdA gene is restricted to serotypes of Salmonella enterica subspecies I and contributes to efficient and prolonged fecal shedding. Infect Immun 2000; 68: 2720-7.

44 Kingsley RA, Santos RL, Keestra AM, Adams LG, Baumler AJ. Salmonella enterica serotype Typhimurium ShdA is an outer membrane fibronectin-binding protein that is expressed in the intestine. Mol Microbiol 2002; 43: 895-905.

45 Kingsley RA, Abi Ghanem D, Puebla-Osorio N, Keestra AM, Berghman L, Baumler AJ. Fibronectin binding to the Salmonella enterica serotype Typhimurium ShdA autotransporter protein is inhibited by a monoclonal antibody recognizing the A3 repeat. J Bacteriol 2004; 186: 4931-9.