Letter

Comparative genome analysis of *Salmonella* Enteritidis PT4 and *Salmonella* Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways

Nicholas R. Thomson,[1,9] Debra J. Clayton,[2] Daniel Windhorst,[3] Georgios Vernikos,[1] Susanne Davidson,[2] Carol Churcher,[1] Michael A. Quail,[1] Mark Stevens,[2] Michael A. Jones,[4] Michael Watson,[2] Andy Barron,[1] Abigail Layton,[2] Derek Pickard,[1] Robert A. Kingsley,[1] Alex Bignell,[1] Louise Clark,[1] Barbara Harris,[1] Doug Ormond,[1] Zahra Abdellah,[1] Karen Brooks,[1] Inna Cherevach,[1] Tracey Chillingworth,[1] John Woodward,[1] Halina Norberczak,[1] Angela Lord,[1] Claire Arrowsmith,[1] Kay Jagels,[1] Sharon Moule,[1] Karen Mungall,[1] Mandy Sanders,[1] Sally Whitehead,[1] Jose A. Chabalgoity,[5] Duncan Maskell,[6] Tom Humphrey,[7] Mark Roberts,[8] Paul A. Barrow,[4] Gordon Dougan,[1] and Julian Parkhill[1]

[1]The Pathogen Sequencing Unit, The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, United Kingdom; [2]Division of Microbiology, Institute for Animal Health, Compton, Berkshire RG20 7NN, United Kingdom; [3]Lohmann Animal Health GmbH & Co. KG, 27472 Cuxhaven, Germany; [4]School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington, Leicestershire LE12 5RD, United Kingdom; [5]Department of Biotechnology, School of Medicine, Universidad de la Republica, Montevideo CP 11600, Uruguay; [6]Department of Veterinary Medicine, University of Cambridge, Cambridge CB3 0ES, United Kingdom; [7]School of Clinical Veterinary Science, University of Bristol, Langford, Bristol BS40 5DU, United Kingdom; [8]Institute of Comparative Medicine, Faculty of Veterinary Medicine, University of Glasgow, Glasgow G61 1QH, United Kingdom

We have determined the complete genome sequences of a host-promiscuous *Salmonella enterica* serovar Enteritidis PT4 isolate P125109 and a chicken-restricted *Salmonella enterica* serovar Gallinarum isolate 287/91. Genome comparisons between these and other *Salmonella* isolates indicate that *S.* Gallinarum 287/91 is a recently evolved descendent of *S.* Enteritidis. Significantly, the genome of *S.* Gallinarum has undergone extensive degradation through deletion and pseudogene formation. Comparison of the pseudogenes in *S.* Gallinarum with those identified previously in other host-adapted bacteria reveals the loss of many common functional traits and provides insights into possible mechanisms of host and tissue adaptation. We propose that experimental analysis in chickens and mice of *S.* Enteritidis–harboring mutations in functional homologs of the pseudogenes present in *S.* Gallinarum could provide an experimentally tractable route toward unraveling the genetic basis of host adaptation in *S. enterica*.

[Supplemental material is available online at www.genome.org. The genome sequence data from this study have been submitted to EMBL under accession nos. AM933172 and AM933173.]

Zoonotic pathogens, particularly those associated with veterinary animals in the human food chain, are some of the most important causes of infectious diseases in humans. Pathogens associated with zoonotic infections exhibit a promiscuous phenotype in that they maintain the ability to colonize and potentially cause infections in more than one host species. In contrast, some pathogenic agents are significantly host restricted, or adapted, and are normally only able to cause disease in one host. *Salmonella enterica* is a single bacterial species that includes examples of both promiscuous and host-adapted pathotypes. Isolates from serovars such as *S. enterica* serovar Typhimurium and *S.* Enteritidis predominantly retain the ability to infect more than one mammalian host, including humans, whereas serovars such as *S. enterica* serovars Typhi and *S.* Gallinarum are restricted to humans and chickens, respectively. The ability to transmit between and within particular host populations is centrally important in dictating the epidemiology of infections and the emergence of new diseases.

Before the mid-1980s, *S.* Enteritidis was regarded as an *S. enterica* serovar of minor public health significance, but subsequently this serovar became dominant in terms of human food poisoning in many parts of the world (Rodrigue et al. 1990). National and international legislation regarding the reporting of disease incidence, improved hygiene and biosecurity (Barrow

[9]Corresponding author.
E-mail nrt@sanger.ac.uk; fax 44-(0)-1223-494919.
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.077404.108.

18:1624–1637 ©2008 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/08; www.genome.org

2000), and vaccination have contributed to controlling *S.* Enteritidis levels in poultry and consequently in man in Europe, but levels of infection remain significant. Most recent isolates of *S.* Enteritidis are regarded as promiscuous in the sense that they can cause infections in mice, retain the ability to colonize the tissues of chickens, and cause gastroenteritis in man.

S. Gallinarum, the causative agent of fowl typhoid, is a predominantly avian-restricted serovar (Shivaprasad 2000). Interestingly, in common with the human-restricted serovar *S.* Typhi, the chicken-adapted *S.* Gallinarum causes an invasive typhoid-like disease. Thus, here host adaptation appears to have coevolved with loss of the intestinal lifestyle and the acquisition of the ability to cause systemic infection. *S.* Gallinarum still causes a disease of worldwide economic significance, and although it has been largely controlled in countries with strong health control policies, largely through serology-based test and slaughter schemes, it remains a problem elsewhere. Multi locus enzyme electrophoresis analyses of isolates of *S.* Enteritidis and *S.* Gallinarum indicate that, together with isolates of *S.* Dublin and *S.* Pullorum, they form a related strain cluster that share the same lipopolysaccharide-based O structure (O-1, 9, 12 characteristic of serogroup D). The nonmotile *S.* Gallinarum and *S.* Pullorum were previously suggested to have split independently from a motile ancestor related to *S.* Enteritidis (Li et al. 1993; McMeechan et al. 2005). Nonmotility in *S.* Gallinarum has been partially attributed to mutations in the flagellin subunit gene *fliC* gene (Kilger and Grimont 1993), which would normally express the phase 1 g, m antigens characteristic of *S.* Enteritidis. Nonmotility may enhance the ability to invade systemically from the gut by avoiding the TLR-5-induced pro-inflammatory responses of the host (Kaiser et al. 2000; Iqbal et al. 2005).

Here we report the full genome sequences of representative isolates of *S.* Enteritidis and *S.* Gallinarum and provide a detailed comparative genomic analysis of the two serovars. These data have been used to provide insight into the biology, mechanisms of host/tissue adaptation, and evolutionary relationships of these important pathogens.

Results and Discussion

General features of the *S.* Enteritidis PT4 strain P125109 and *S.* Gallinarum strain 287/91 genomes

The complete genome sequences of the promiscuous *S.* Enteritidis PT4 strain P125109 (hereafter *S.* Enteritidis PT4; EMBL accession no. AM933172) and the highly host-adapted chicken pathogen *S.* Gallinarum strain 287/91 (hereafter *S.* Gallinarum 287/91; EMBL accession no. AM933173) were determined and annotated. The main features are summarized in Table 1 and Figure 1, where they are compared with *S.* Typhimurium strain LT2 (hereafter *S.* Typhimurium LT2) (McClelland et al. 2001). The most striking feature of the analysis is the predominant similarity and synteny of core regions of the genomes, including many of the *Salmonella* pathogenicity islands (SPI). Indeed, this comparative analysis highlights an extremely close relationship between the genomes of *S.* Enteritidis and *S.* Gallinarum, suggesting the latter is a direct evolutionary descendent of the former. However, in comparison to *S.* Enteritidis PT4, *S.* Gallinarum 287/91 harbors a significantly higher number of predicted pseudogenes. Although the number of pseudogenes in *S.* Enteritidis PT4 is slightly higher than reported for *S.* Typhimurium LT2, it is in line with levels described in other broad host range enteric pathogens such as *Yersinia enterocolitica* (Thomson et al. 2006). In contrast, the number of pseudogenes in *S.* Gallinarum 287/91 is closer to that of the human-restricted *S.* Typhi CT18 (204 pseudogenes) (Parkhill et al. 2001) and *S. enterica* serovar Paratyphi A (173 pseudogenes) (McClelland et al. 2004).

Whole-genome comparisons of *S.* Enteritidis PT4 and *S.* Typhimurium LT2

Initially, the genome of *S.* Enteritidis PT4 was compared with that of *S.* Typhimurium LT2, a well-characterized and fully sequenced *S. enterica* isolate. *S.* Enteritidis PT4 and *S.* Typhimurium LT2 are both representatives of serovars able to cause enteritis in a broad range of hosts and produce murine typhoid, but they also show significant phenotypic differences, including serovar type. An alignment of the genome of *S.* Enteritidis PT4 with that of *S.* Typhimurium LT2 revealed colinearity except for an inversion about the terminus in *S.* Typhimurium LT2 (Fig. 1) (McClelland et al. 2001), with >90% of coding sequences (CDS) forming an extensive core gene-set (Figs. 2, 3). The average nucleotide identity between the shared orthologs is 98.98% compared with 99.7% between those of LT2 and a second fully sequenced *S.* Typhimurium strain SL1344 (data not shown). The genes that are only present either in *S.* Enteritidis PT4 or *S.* Typhimurium LT2 form 6.4% and 9.6% of their respective genomes (Fig. 3). The majority of *S.* Enteritidis PT4 unique CDS are in clusters from >3 kb up to >40 kb, but there are very few indels of <3 kb (Fig. 2; Table 2). We refer to these nonshared gene clusters as regions of difference (ROD). CDS present in *S.* Enteritidis PT4 but absent from *S.* Typhimurium LT2 are dominated by prophage-related functions, although other functional classes are represented (Fig. 3).

Table 1. General properties of *S. enterica* serovar genomes

	S. enterica serovars			
Serovar	**Enteritidis**	**Typhimurium**	**Gallinarum**	**Typhi**
Strain	P125109 (PT4)	LT2	287/91	CT18
Size	4,685,848	4,857,432	4,658,697	4,809,037
Percent G + C content (%)	52.17	52.22	52.20	52.09
No. of CDS	4318	4451	4274	4599
Coding density	85.5%	86.8%	79.9%	87.6%
Average gene size	953	947	939	958
rRNA operons	7	7	7	7
tRNA	84	85	75	78
Pseudogenes[a]	113	25	309	204

[a]Taken from original publications (see text).

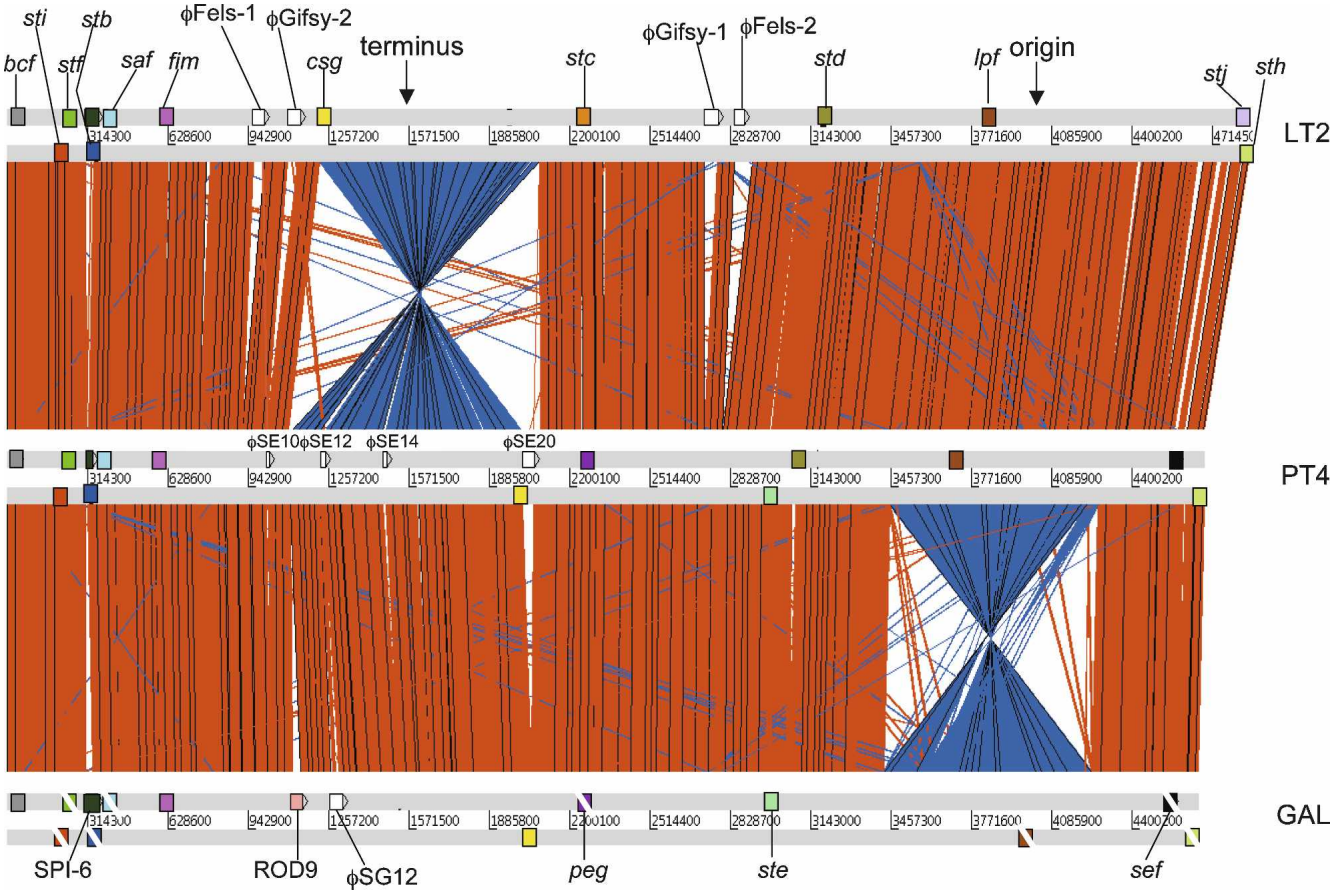

Figure 1. Global comparison between *S.* Typhimurium, *S.* Enteritidis, and *S.* Gallinarum. ACT comparison (http://www.sanger.ac.uk/Software/ACT) of amino acid matches between the complete six-frame translations (computed using TBLASTX) of the whole-genome sequences of *S.* Typhimurium LT2 (LT2), *S.* Enteritidis PT4 (PT4), and *S.* Gallinarum 287/91 (GAL). Forward and reverse strands of DNA are shown for each genome (light gray horizontal bars). The red bars between the DNA lines represent individual TBLASTX matches, with inverted matches colored blue. The position of all the fimbrial operons in these three genomes are marked as colored boxes positioned on the forward and reverse strands of DNA. Analogous fimbrial operons are colored the same. The boxes of fimbrial operons that include pseudogenes are crossed with a white line. Other genomic features are only shown if they constitute breaks in synteny between genomes. The position of the origin and terminus are marked (solid black arrows).

Gene sets common to both *S.* Enteritidis PT4 and *S.* Typhimurium LT2

Within the core genes, there are many of the functions associated with virulence and host interactions and include SPIs and fimbrial operons. With the exception of *SPI-6*, *SPI-9*, and *SPI-10*, the other SPIs in *S.* Enteritidis PT4 are closely related to their equivalents in *S.* Typhimurium LT2 (Fig. 2) (McClelland et al. 2001). Of the three SPIs that vary, *S.* Enteritidis PT4 *SPI-10* only encodes the *sef* fimbrial operon, consistent with this region being mosaic in isolates of different serovars (Edwards et al. 2000; Collighan and Woodward 2001; Bishop et al. 2005). The *SPI-9* CDS *SE2609*, encoding a large repetitive exported protein, appears intact in *S.* Enteritidis PT4 unlike the ortholog, *STM2689*, in *S.* Typhimurium LT2. The *SPI-6* region of *S.* Enteritidis PT4 is 22 kb in size compared with 47 kb in *S.* Typhimurium LT2. *SPI-6* varies markedly in size in all the other sequenced *Salmonella*, including *S.* Typhi (McClelland et al. 2001, 2004; Parkhill et al. 2001; Chiu et al. 2005). Of the other known SPIs, *SPI-8*, *SPI-7*, and *SPI-15* are absent from *S.* Enteritidis PT4 (Parkhill et al. 2001; Vernikos and Parkhill 2006). Conversely, *SPI-17* is present in *S.* Enteritidis PT4 but absent from *S.* Typhimurium LT2. The *S.* Enteritidis PT4 *SPI-17* is a degenerate prophage encoding CDS known to be involved in O-antigen conversion in other systems (Vernikos and Parkhill 2006).

S. Enteritidis PT4 harbors 13 fimbrial clusters, 10 of which are highly conserved in *S.* Typhimurium LT2 with orthologous genes sharing >97% nucleotide identity and inserted at the same sites in both genomes (Fig. 1; Table 2). The only exceptions to this are *safA*, *safB*, and *stdA*, where the *S.* Enteritidis PT4 and *S.* Typhimurium LT2 orthologs show 81%, 87%, and 89% nucleotide identity, respectively. The *S.* Enteritidis PT4 fimbrial clusters not found in *S.* Typhimurium LT2 include a novel cluster we have termed *peg*, which is inserted at the same location as the *S.* Typhimurium LT2 *stc* operon and is so far restricted to *S.* Enteritidis, *S.* Gallinarum 287/91, and *S.* Paratyphi A. The *peg* fimbrial proteins show 58%–64% identity with their predicted functional equivalents in the *S.* Typhimurium LT2 *stc* cluster (Table 2). Of the remaining fimbrial clusters, *ste* is absent from *S.* Typhimurium LT2, but there is a deletion remnant of the *ste* major pilin subunit remaining at the analogous site (*S.* Typhimurium LT2; positions 3,102,016–3,102,150 bps). Fimbrial operon *stj* is present in *S.* Typhimurium LT2 and replaces a gene of unknown function still present in *S.* Enteritidis PT4 (*SEN4331A*). Thus, in common

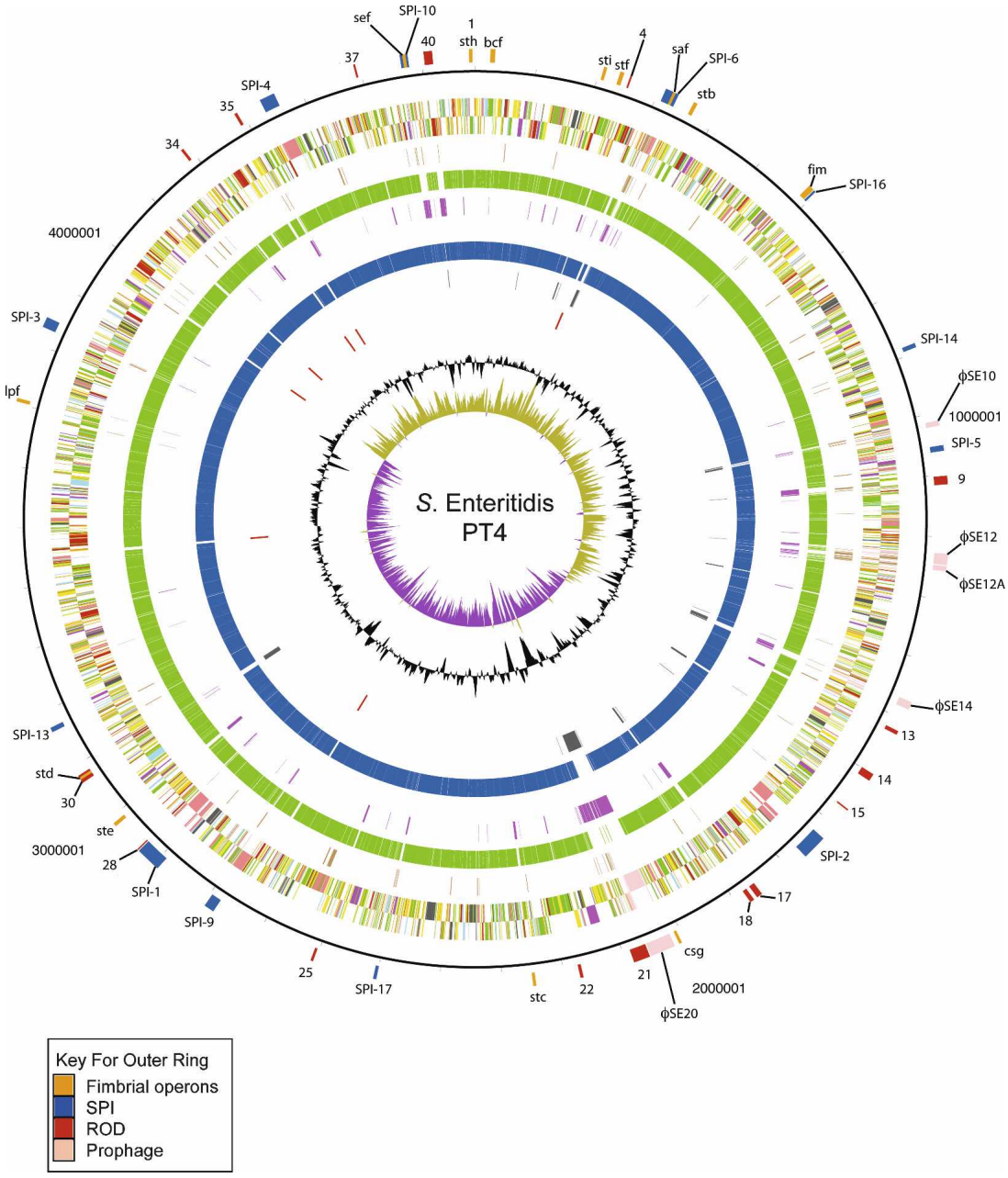

Figure 2. Circular representation of the *S.* Enteritidis PT4 chromosome. From the outside in, the *outer* circle 1 marks the position of regions of difference (mentioned in the text) and is detailed in Table 2. Circle 2 shows the size in base pairs. Circles 3 and 4 show the position of CDS transcribed in a clockwise and anti-clockwise direction, respectively (for color codes see below); circle 5 shows the position of *S.* Enteritidis PT4 pseudogenes. Circles 6 and 8 show the position of *S.* Enteritidis PT4 genes that have orthologs (by reciprocal FASTA analysis) in *S.* Typhimurium strain LT2 (all CDS colored green) and *S.* Gallinarum strain 287/91 (all CDS colored blue), respectively. Circles 7 and 9 show the position of *S.* Enteritidis PT4 genes that lack orthologs in (by reciprocal FASTA analysis) in *S.* Typhimurium strain LT2 (all CDS colored pink) and *S.* Gallinarum strain 287/91 (all CDS colored gray), respectively. Circle 10 shows the position of *S.* Enteritidis PT4 rRNA operons (red). Circle 11 shows a plot of G + C content (in a 10-kb window). Circle 12 shows a plot of GC skew ([G − C]/[G + C]; in a 10-kb window). Genes in circles 3 and 4 are color-coded according to the function of their gene products: dark green, membrane or surface structures; yellow, central or intermediary metabolism; cyan, degradation of macromolecules; red, information transfer/cell division; cerise, degradation of small molecules; pale blue, regulators; salmon pink, pathogenicity or adaptation; black, energy metabolism; orange, conserved hypothetical; pale green, unknown; and brown, pseudogenes.

with other promiscuous salmonellae, *S.* Enteritidis PT4 harbors multiple functional fimbrial operons (Townsend et al. 2001).

In addition to the gene remnants found for the *ste* fimbrial operon *S.* Enteritidis PT4, RODs *ROD17*, *ROD25*, *ROD34*, *ROD35*, and *ROD37* (Table 2) that fall outside of the core gene-set also have discernable remnants in *S.* Typhimurium LT2 or are conserved in other *S. enterica* and so are likely to have been present in a precursor of *S.* Enteritidis PT4, shared with *S.* Typhimurium LT2, and subsequently deleted from *S.* Typhimurium LT2.

Gene sets only present in *S.* Enteritidis PT4 but not *S.* Typhimurium LT2

By analyzing the genetic context of the *S.* Enteritidis PT4 specific RODs, we distinguish between those likely to have been acquired independently from those that may have been deleted from *S.* Typhimurium LT2 (discussed above). Examples of likely acquisitions include *SPI-17*, *ROD4*, *ROD9*, *ROD13*, *ROD21*, *ROD22*, *ROD28*, *ROD40*, ϕSE14, and ϕSE20 (Table 2). RODs unique to *S.* Enteritidis PT4 include potentially mobile genomic islands, clusters of genes encoding metabolic functions and prophage-like elements, as well as a variable assortment of fimbrial operons already described (summarized in Table 2).

Genomic islands

ROD21 is the only *S.* Enteritidis PT4 genomic island not found in *S.* Typhimurium LT2 and has features characteristic of mobile genetic elements (Table 2). *ROD21* shares significant structural conservation with conserved genomic loci present in a range of bacteria. These islands all display an unusual G + C profile, whereby regions conserved between islands show a higher G + C content compared with the variable island-specific regions (Fig. 4) (Williamson and Free 2005; this study). Surprisingly, most of the *ROD21*-related islands encode paralogs of H-NS (*hnsB*) and/or an H-NS antagonist, *hnsT* (Williamson and Free 2005; Navarre et al. 2006; Doyle et al. 2007). These paralogs may play a role in relieving any potential fitness burden associated with sequestering H-NS to these low G + C DNA elements (Doyle et al. 2007).

Of the other *S.* Enteritidis PT4-specific RODs, *ROD13* encodes five CDS displaying sequence similarities and synteny with genes associated with the uptake and catabolism of the hexonate sugar acid *L*-idonate encoded by the *gntII* locus of *Escherichia coli* (Table 2) (Bausch et al. 1998). Although the substrate for this system is unclear, it is known that colonic mucus contains several sugar acids that represent an important source of nutrients and that *E. coli* mutants unable to utilize them are unable to colonize the mouse large intestine (Sweeney et al. 1996). Moreover, genes involved in the transport of gluconate and related hexonates are up-regulated in *S.* Typhimurium in macrophage, suggesting that they may also be an important source of carbon for intracellular bacteria (Eriksson et al. 2003).

The *S.* Enteritidis PT4 RODs also include loci that are highly variable in the salmonellae and, for *ROD40*, between the wider *Enterobacteriaceae*. *ROD40* locus encodes a Type I restriction/modification system and is analogous to the variable *E. coli* immigration control region (ICR) (Raleigh 1992; Titheradge et al. 1996).

Prophage

Prophage are known to drive diversity in *S. enterica*, and thus, it is not surprising that many *S.* Enteritidis PT4–specific RODs are prophage-like elements, including ϕSE10, ϕSE12, ϕSE12A, ϕSE14, and ϕSE20 (Fig. 2; Table 2) (Thomson et al. 2004; Cooke et al. 2007). All these prophage regions are related and carry the same cargo genes as prophage found previously in other *S. enterica* (see Table 2), including genes encoding type three secretion system (TTSS) effector proteins—*sseK3*, *sspH2*, *gogA*, *sseI*, and *sopE*; the PhoPQ-activated genes *pagK* and *pagM*; as well as *sodCI* encoding a Cu/Zn superoxide dismutase known to be an important colonization factor for *S.* Typhimurium (Stanley et al. 2000; Figueroa-Bossi et al. 2001; Mmolawa et al. 2003; Thomson et al. 2004). Of the six prophage-related regions, only ϕSE20 appears intact and probably represents a recent insertion event, whereas remnants of ϕSE12A are also present at the same location in *S.* Typhimurium LT2 and probably represent the most ancient phage insertion that has been maintained in these two *Salmonella* lineages. However, the number of remnants and intact cargo genes found on the *S.* Enteritidis PT4 prophage highlights the importance of these elements for gene sampling and increasing the overall diversity and even pathogenic potential of salmonellae.

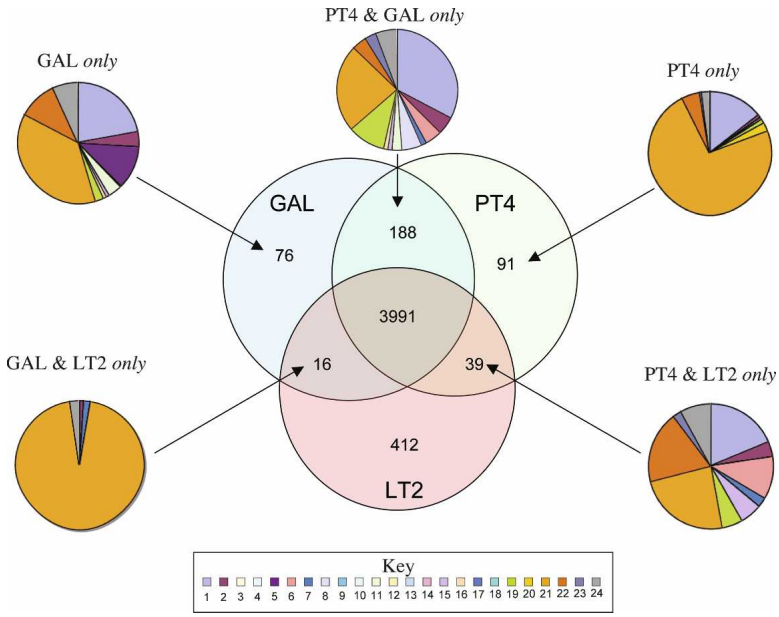

Figure 3. Distribution of orthologous CDS in *S.* Enteritidis PT4, *S.* Typhimurium LT2, and *S.* Gallinarum 287/91. The Venn diagram shows the number of genes unique or shared between two other *S. enterica* serovars (see Methods). The associated pie charts show the breakdown of the functional groups assigned for CDS in relevant sections of the Venn diagram. Color code for the pie charts is as follows: (1) hypothetical proteins, (2) conserved hypothetical proteins, (3) chemotaxis and motility, (4) chromosomal replication, (5) chaperones, (6) protective responses, (7) transport and binding proteins, (8) adaptations to atypical conditions, (9) cell division, (10) macromolecule degradation, (11) synthesis and modification of macromolecules, (12) amino acid biosynthesis, (13) biosynthesis of cofactors, prosthetic groups and carriers, (14) central intermediary metabolism, (15) small-molecule degradation, (16) energy metabolism, (17) fatty acid biosynthesis, (18) nucleosides and nucleotide biosynthesis and metabolism, (19) periplasmic/exported/lipoproteins, (20) ribosomal proteins, (21) laterally acquired (including prophage CDS), (22) pathogenicity and virulence, (23) general regulation, and (24) miscellaneous function. PT4 indicates *S.* Enteritidis PT4; LT2, *S.* Typhimurium LT2; and GAL, *S.* Gallinarum 287/91

Whole-genome sequence of *S.* Gallinarum 287/91 and comparisons with *S.* Enteritidis PT4 and *S.* Typhimurium LT2

One of the striking features of the *S.* Gallinarum 287/91 genome is the high simi

Table 2. The *S.* Enteritidis PT4 variable genome regions identified by genome sequencing

Label[a]	PT4 CDS range	GAL CDS range	Locus name(s)[b]	Size in PT4[c]	General description of locus	Where present
bcf	*SEN0020–SEN0027*	*SG0023–SG0030*	*bcf*	7.6 kb	Fimbrial operon (bovine colonization factor)	GAL PT4 LT2
sti	*SEN0179–SEN0182*	*SG0177–SG0180*	*sti*	4.9 kb	Fimbrial operon (*S.* Typhi I)	GAL PT4 LT2
stf	*SEN0200–SEN0205*	*SG0199–SG0204*	*stf*	6.4 kb	Fimbrial operon (*S.* Typhi F)	GAL PT4 LT2
4	*SEN0216*	*NP*	*ROD4*	3 kb	Viral enhancing factor (metalloprotease)	PT4, NI LT2, or GAL
SPI-6	*SEN0267–SEN0290*	*SG0263–SG0318*	*SPI-6*	17.6 kb (44 kb)	*Salmonella* pathogenicity island (tRNA-*asp*)	GAL PT4 LT2. Variable in all three (Parkhill et al. 2001)
saf	*SEN0281–SEN0284*	*SG0308–SG0312*	*saf*	4.3 kb	Part of SPI-6. Fimbriae: *Salmonella* atypical fimbriae	GAL PT4 LT2
stb	*SEN0319–SEN0323*	*SG0346–SG0350*	*stb*	5.9 kb	Fimbrial operon (*S.* Typhi B)	GAL PT4 LT2
fim	*SEN0524–SEN0533*	*SG0555–SG0564*	*fim*	9 kb	Fimbrial operon (tRNA-*arg*)	GAL PT4 LT2
SPI-16	*SEN0535–SEN0537*	*SG0567–SG0569*	*SPI-16*	3.3 kb	Phage remnant. Cargo: LPS modification genes (tRNA-*arg*)	GAL PT4 LT2 (Vernikos and Parkhill 2006)
SPI-14	*SEN0800–SEN0805*	*SG0835–SG0840*	*SPI-14*	6.8 kb	Electron transfer and regulatory CDS	GAL PT4 LT2 (Shah et al. 2005)
ϕSE10	*SEN0908A–SEN0921*	*NP*	ϕSE10	8.2 kb	Prophage (remnant). Related to ϕGifsy-2. Cargo: *sseI*, *gtgE*, *gtgF*	PT4 LT2, NI: GAL
SPI-5	*SEN0951–SEN0958*	*SG0976–SG0985*	*SPI-5*	6.6 kb	*Salmonella* pathogenicity island (tRNA-*ser*)	GAL PT4 LT2 (Wood et al. 1998)
9	*SEN0995–SEN1013B*	*SG1023–SG1058*	*ROD9*	13.7 kb (42 kb)	RHS element, exported proteins and an IcmF-like CDS	GAL. NI: LT2. Degenerate in PT4
ϕSE12	*SEN1131–SEN1156*	*SG1180–SG1234*	ϕSE12 [ϕSG12]	18 kb (45 kb)	Prophage (remnant). Related to ϕGifsy-2 Cargo: *sodCI*, *ompX**, *sopE*, *gogA**, *hokW*	GAL. NI: LT2. Degenerate in PT4
ϕSE12A	*SEN1158–SEN1171B*	*SG1236–SG1249*	ϕSE12A [ϕSG12A]	8 kb	Prophage (remnant). Related to ϕGifsy-2 Cargo: *pagK*, *pagM**	GAL PT4 LT2
ϕSE14	*SEN1378–SEN1398*	*NP*	ϕSE14	12.6 kb	Prophage (remnant). Related to *S.* Typhi ϕST18	PT4, NI: LT2 GAL
13	*SEN1432–SEN1436*	*SG1499–SG1503*	*ROD13*	6 kb	Possible hexonate uptake and catabolism operon	GAL PT4, NI LT2
14	*SEN1499–SEN1509*	*SG1573*–SG1576**	*ROD14*	11.7 kb (1.8 kb)	Drug efflux system, *pqaA*	GAL PT4 LT2. Variable in all three
15	*SEN1565–SEN1567*	*SG1636**	*ROD15*	2.2 kb (0.6 kb)	Exported protease, *mdtI* and *mdtJ* (Nal^r, Fos^r, Det^r) (Nishino and Yamaguchi 2001)	PT4 LT2. Remnant in GAL
SPI-2	*SEN1623–SEN1666*	*SG1694–SG1738*	*SPI-2*	39.8 kb	*Salmonella* pathogenicity island (tRNA-*val*)	PT4 GAL LT2 (Hensel 2000)
17	*SEN1751–SEN1758*	*SG1824–SG1832*	*ROD17*	9.3 kb	Putative membrane-transport system	PT4 GAL. Remnant in LT2 (1,370,980–1,371,012)
18	*SEN1766–SEN1769*	*SG1841**	*ROD18*	6.9 kb (0.3 kb)	*mipA*—outer-membrane scaffold	PT4 LT2. Remnant in GAL
csg	*SEN1903–SEN1909*	*SG1977–SG1983*	*csg*	4.4 kb	Curli fimbrial operon	PT4 GAL LT2
ϕSE20	*SEN1919A–SEN1966*	*NP*	ϕSE20	40.6 kb	Prophage. Related to ϕST64B. Cargo: *sopE**, *sspH1**, *orgA* (tRNA-*ser*)	PT4, NI: GAL LT2
21	*SEN1970–SEN1999*	*SG1996–SG2025*	*ROD21*	26.5 kb	Genomic island (61 bp DR; tRNA-*asn*; 37.5% G + C) Cargo: *hnsT*, *hnsB*	PT4 GAL, NI: LT2
22	*SEN2085A–SEN2085D*	*SG2117–SG2120*	*ROD22*	4.9 kb	Group D LPS O-chain genes *rfbE* and *rfbS*	PT4 GAL. Different genes in LT2 (group B O-chain)
peg	*SEN2144A–SEN2145B*	*SG2182–SG2186*	*peg*	4.8 kb	Fimbrial operon	PT4 GAL, NI: LT2
SPI-17	*SEN2375A–SEN2380A*	*SG2425–SG2430A*	*SPI-17*	3.6 kb	Prophage remnant. Cargo: *gtrA*, *gtrB** *gtrC* (tRNA-*arg*)	PT4 GAL, NI: LT2 (Vernikos and Parkhill 2006)
25	*SEN2471–SEN2473*	*SG2522–SG2524*	*ROD25*	4 kb	Membrane transport system	PT4 GAL. Remnant in LT2 (STM2492*)
SPI-9	*SEN2609–SEN2612*	*SG2666–SG2671*	*SPI-9*	16.3 kb	*Salmonella* pathogenicity island (10Sa_RNA)	PT4 GAL LT2 (Parkhill et al. 2001)
SPI-1	*SEN2703–SEN2744*	*SG2764–SG2806*	*SPI-1*	40.2 kb	*Salmonella* pathogenicity island	PT4 GAL LT2 (Hansen-Wester and Hensel 2001)
28	*SEN2746–SEN2746A*	*SG2808–SG2811*	*ROD28*	2 kb	Membrane proteins	PT4 GAL, NI: LT2

(continued)

Table 2. *Continued*

Label[a]	PT4 CDS range	GAL CDS range	Locus name(s)[b]	Size in PT4[c]	General description of locus	Where present
ste	*SEN2794–SEN2799*	*SG2859–SG2864*	*ste*	5.6 kb	Fimbrial operon (*S.* Typhi E)	PT4 GAL. Remnant in LT2 (major pilin remnant: 3,102,016–3,102,150)
30	*SEN2864–SEN2878*	*SG2930**	*ROD30*	13 kb (0.16 kb)	*rcnA*, the *std* fimbrial, operon	PT4 LT2. Remnant in GAL
std	*SEN2871–SEN2873*	*NP*	*std*	5.0 kb	Within *ROD30*. Fimbrial operon (*S.* Typhi D)	PT4 LT2, NI: GAL
SPI-13	*SEN2960–SEN2966*	*SG3011–SG3017*	*SPI-13*	7.4 kb	*Salmonella* pathogenicity island (tRNA-*phe*)	PT4 GAL LT2 (Shah et al. 2005)
lpf	*SEN3459–SEN3463*	*SG3793–SG3798*	*lpf*	5.5 kb	Fimbrial operon (long polar fimbriae)	PT4 GAL LT2
SPI-3	*SEN3572–SEN3586*	*SG3665–SG3680*	*SPI-3*	16.6 kb	*Salmonella* pathogenicity island (tRNA-*selC*)	PT4 GAL LT2 (Blanc-Potard et al. 1999)
34	*SEN3896–SEN3898*	*SG3312–SG3314*	*ROD34*	4.2 kb	Amino acid metabolic CDS	PT4 GAL, NI: LT2
35	*SEN3978–SEN3981*	*SG4054–SG4058*	*ROD35*	4.5 kb	Unknown	PT4 GAL, NI: LT2
SPI-4	*SEN4026–SEN4032*	*SG4100–SG4105*	*SPI-4*	25 kb	*Salmonella* pathogenicity island	PT4 GAL LT2 (Wong et al. 1998; Parkhill et al. 2001)
37	*SEN4165–SEN4166*	*SG4241–SG4242*	*ROD37*	3 kb	Unknown	PT4 GAL, NI: LT2
SPI-10	*SEN4244–SEN4254*	*SG4311–SG4325*	*SPI-10*	10 kb	*Salmonella* pathogenicity island (tRNA-*leu*)	GAL PT4, NI: LT2 (Parkhill et al. 2001)
sef	*SEN4247–SEN4251*	*SG4318–SG4322*	*sef*	5.1 kb	Fimbrial operon (*Salmonella* Enteritidis fimbriae)	PT4 GAL, NI: LT2
40	*SEN4283–SEN4292*	*SG4349–SG4357*	*ROD40*	13.5 kb (10 kb)	Type I restriction-modification system	PT4. Degenerate in GAL, NI: LT2
sth	*SEN4347–SEN4351*	*SG4413–SG4417*	*sth*	5.5 kb	Fimbrial operon (*S.* Typhi H)	PT4 GAL LT2
42	*SEN3843*–SEN3844**	*SG3367–SC3368*	*ROD42*	1 kb	C4-dicarboxylate transporters	GAL LT2. Deleted from PT4

[a]Labels used to mark these regions on the outer ring of Figures 2 and 5.
[b]Square brackets indicate the name in *S.* Gallinarum 287/91.
[c]Numbers in parentheses indicate the size of the analogous region in *S.* Gallinarum. Only shown when significantly different from that found in *S.* Enteritidis PT4.
(*) Gene remnant; (DR) direct repeat; (NI) not detected in; (LPS) lipopolysaccharide locus; (LT2) *S.* Typhimurium LT2; (GAL) *S.* Gallinarum 287/91; (PT4) *S.* Enteritidis PT4; (Nal^r) nalidixic acid resistance; (Fos^r) fosfomycin resistance; (Det^r) detergent resistance.

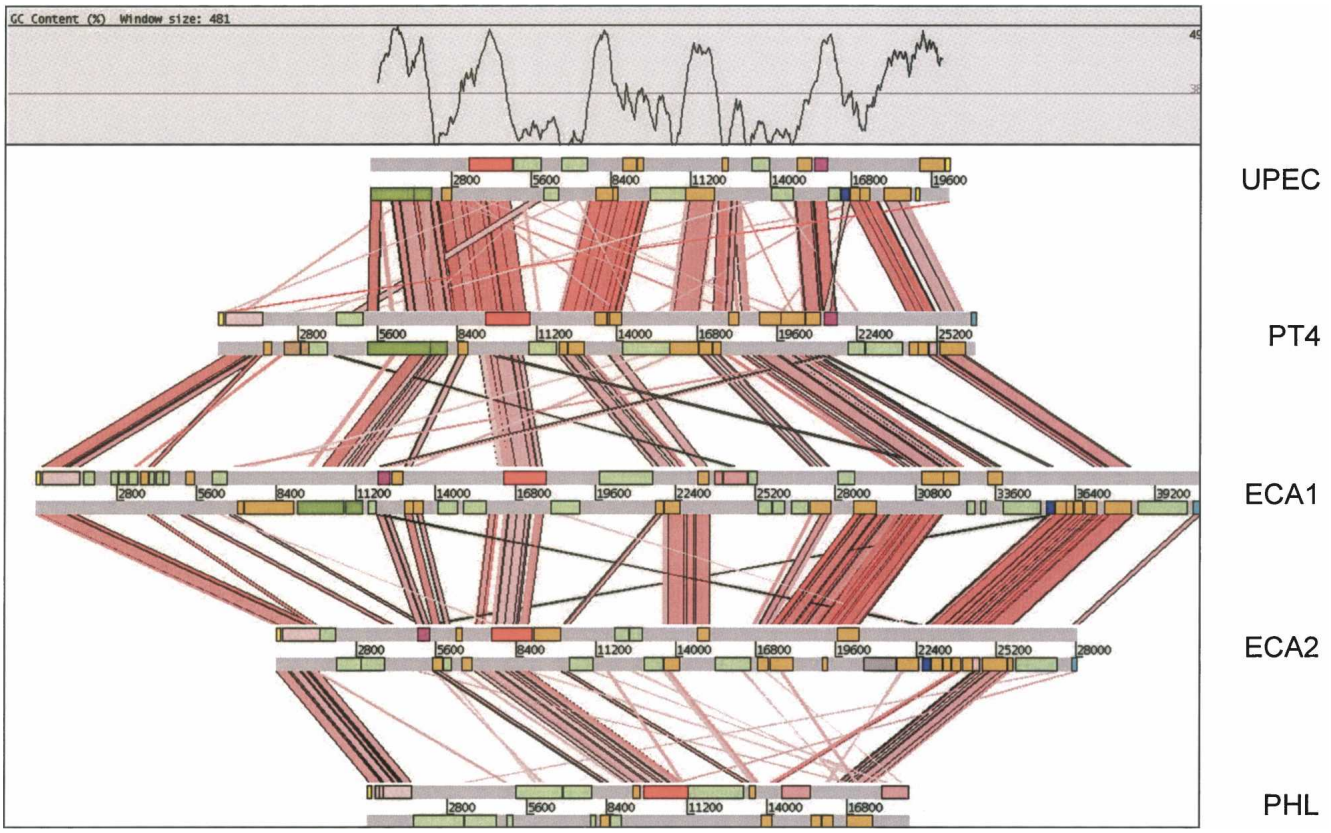

Figure 4. Comparison of the *ROD21* locus of *S.* Enteritidis with related genomic islands. ACT comparison (http://www.sanger.ac.uk/Software/ACT) of amino acid matches between the complete six-frame translations (computed using TBLASTX) of *ROD21* compared with related loci uropathogenic *E. coli* strain CFT073 (UPEC), *Erwinia carotovora* sbsp. *atroseptica* strain SCRI1043 (two loci: *ECA1* and *ECA2*), and *Photorhabidus luminescens*,sbsp. *laumondii* TT01 (PHL; see Methods). The red bars spanning between the genomes represent individual TBLASTX matches. CDS are marked as colored boxes positioned on the horizontal gray DNA bars: (orange) genes conserved in two or more of the genomic islands; (light green) variable genes of unknown function; (dark pink) *hnsB*; (dark blue) *hnsT*; (light pink) integrase; (dark green) type IV pilin-associated genes; (red) plasmid-related mobility functions; (salmon pink) transposase-related genes; (yellow) tRNA genes; (light blue)repeats. The G + C profile for the UPEC loci is shown above. The scale is marked in base pairs.

larity with *S.* Enteritidis PT4 compared with *S.* Typhimurium LT2 (Figs. 3, 5). The average nucleotide identities of orthologs shared between *S.* Gallinarum 287/91 and *S.* Enteritidis PT4 were higher (99.7%) than those found in LT2 (98.93%). Another obvious feature is the massive accumulation of pseudogenes in *S.* Gallinarum 287/91 compared with *S.* Enteritidis PT4 and *S.* Typhimurium LT2. The genome of *S.* Gallinarum 287/91 is slightly smaller than *S.* Enteritidis PT4, carries significantly fewer tRNA genes (Table 1), and is colinear except for a single inversion (817 kb; about the rRNA operons) and translocation of a region (49 kb) located between two different rRNA operons (Figs. 1, 5).

The number of CDS unique to *S.* Gallinarum 287/91 (76) or shared exclusively between *S.* Gallinarum 287/91 and *S.* Typhimurium LT2 (16) was small and predominantly phage-associated (Figs. 3, 5). Moreover, genes from both of these categories all fell within regions (*SPI-6*, *ROD9*, *ROD42*, and ϕSG12) that are present in *S.* Enteritidis PT4 but appear to be in the process of being lost. Consequently, these genes are unlikely to be recent acquisitions by *S.* Gallinarum 287/91.

Of the 130 CDS specific to *S.* Enteritidis PT4, compared with *S.* Gallinarum 287/91, those associated with *ROD4* and prophages ϕSE10, ϕSE14, and ϕSE20 (82 CDS) appear to be recent acquisitions with no evidence of them ever being present in *S.* Gallinarum 287/91(Fig. 3; Table 2). Of the remaining 48 CDS in this category, 21 were located on loci (*ROD15*, *ROD18*, and *ROD30*) (Table 2) almost entirely deleted from *S.* Gallinarum 287/91. The others were located on shared loci such as *ROD14* and *SPI-6* that are degenerate in both serotypes, compared with *S.* Typhimurium LT2. The functions that these RODs encode in *S.* Enteritidis PT4 are summarized in Table 2.

Thus, we provide compelling genetic evidence that *S.* Enteritidis and *S.* Gallinarum are recently diverged clones. On this conclusion, we have plotted the most parsimonious explanation for the observed gene flux following the divergence of *S.* Typhimurium LT2, *S.* Enteritidis PT4, and *S.* Gallinarum 287/91 (Fig. 6).

Functional gene loss and pseudogene formation

In addition to the large scale deletion, there is further evidence of reductive evolution in *S.* Gallinarum 287/91 in the form of 309 putative pseudogenes that carry frameshifts or premature stop codons or that are remnants of genes present in other bacteria. Remarkably, this represents ~7% of the total coding capacity of the genome and includes genes from many functional categories, including metabolism and virulence (for a full list, see Supplemental Table 1).

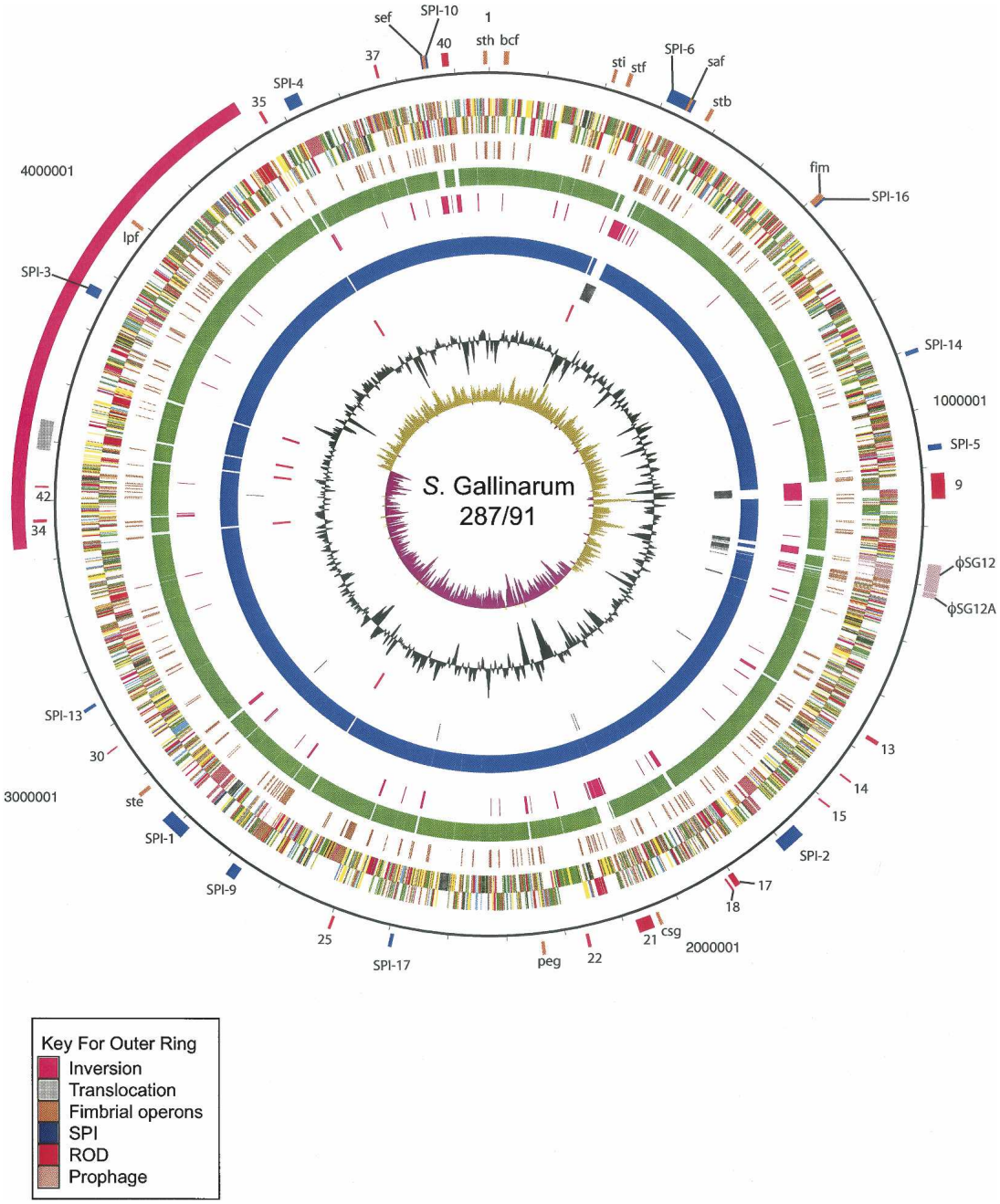

Figure 5. Circular representation of the *S.* Gallinarum 287/91 chromosome. For a full description of this figure, see legend to Figure 2. The exception is that circle 5 shows the position of *S.* Gallinarum 287/91 pseudogenes. Circles 6 and 8 show the position of *S.* Gallinarum 287/91 genes which have orthologs (by reciprocal FASTA analysis) in *S.* Typhimurium LT2 (all CDS colored green) or *S.* Enteritidis PT4 (all CDS colored blue), respectively. Circles 7 and 9 show the position of *S.* Gallinarum 287/91 genes that lack orthologs in (by reciprocal FASTA analysis) in *S.* Typhimurium strain LT2 (all CDS colored pink) and *S.* Enteritidis PT4 (all CDS colored gray), respectively. Circle 10 shows the position of the *S.* Gallinarum 287/91 rRNA operons (red).

S. Gallinarum 287/91 has lost several pathways, which are likely to narrow the spectrum of substrates available for use as sources of carbon and energy. These include *malS* encoding a periplasmic alpha-amylase required for growth on long chain maltodextrins (Schneider et al. 1992), those affecting the catabolism of (*D*)-glucarate (*gudD*) and the mutation in the *hyaF* gene (within the hydrogenase 1 [*hya*] gene cluster).

S. Gallinarum 287/91 also possesses multiple mutations in genes within all three operons required for the breakdown 1,2-propanediol: *ttr*, *cbi*, and *pdu* operons directing tetrathionate respiration; coenzyme B_{12} biosynthesis (B_{12}; cobalamine); and 1,2-propanediol degradation, respectively (Supplemental Table 1) (Roth et al. 1996). 1,2-Propanediol is an important source of energy for *S.* Typhimurium, and *cbi* mutants are significantly attenuated in their ability to grow in macrophages (Klumpp and Fuchs 2007). Consequently, for most of the salmonellae, the ability to degrade propanediol is the likely selective pressure maintaining the *cbi* and *ttr* genes, and the loss of function of any of

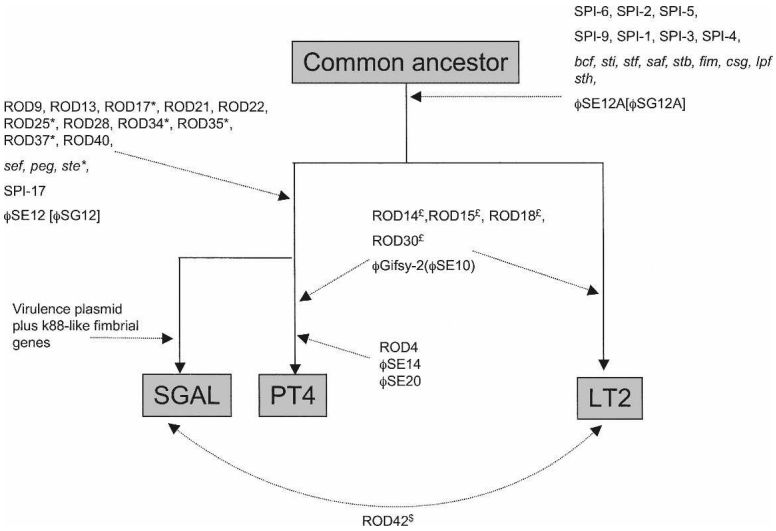

Figure 6. Line diagram to represent the whole-genome differences of *S.* Enteritidis PT4, *S.* Typhimurium strain LT2, and *S.* Gallinarum strain 287/91. A summary of the observed loss and gain of RODs described in Table 2. The diagram is based on the assumption that following the divergence of PT4 and LT2 from a common ancestor PT4 and GAL have subsequently diverged. Branches are not intended to infer phylogenetic distance. Evidence that a locus was once present in LT2, PT4, or GAL (see legend to Figure 1) but has subsequently been deleted from that genome is marked by the suffix *, $, or £, respectively. Brackets indicate the name for locus in GAL. Parentheses indicate the name for locus in PT4. Dotted arrows mark the position in the pseudo-tree at which that ROD(s) appears

these three pathways was probably the precursor to the degeneration of the others in *S.* Gallinarum 287/91. Mutations within the *cbi*, *pdu*, and *ttr* genes was also a feature of the *S.* Typhi CT18 genome (Table 3), suggesting that these mutations may be characteristic of more invasive *Salmonella* serotypes.

Uniquely among the *Salmonella*, *S.* Gallinarum 287/91 and *S.* Pullorum are unable to make glycogen (Supplemental Table 2) (McMeechan et al. 2005). The genome sequence data revealed that mutations in the glycogen biosynthetic pathway are extensive in *S.* Gallinarum 287/91, including *glgA*, *glgB*, and *glgC*, which are responsible for all steps in biosynthesis. Although the significance of these mutations are not clear, they may explain in part the poor survival of this bacterium outside of the host (McMeechan et al. 2005).

Several of the identified *S.* Gallinarum 287/91 pseudogenes lie in amino acid catabolic or biosynthetic pathways. *Salmonella* encodes three pathways for arginine degradation (Reitzer 2003), and *S.* Gallinarum 287/91 carries mutations in two of these: There are deletions or frameshifts in the genes of the arginine *N*-succinyltransferase (AST) (*astA*) and arginine deiminase (ADI) (*arcA*) pathways, respectively. *S.* Gallinarum 287/91 also carries a mutation in *speC* encoding the ornithine decarboxylase, making the one remaining intact arginine catabolic pathway, arginine decarboxylase pathway (ADC), an essential biosynthetic route for putrescine. The mutation in *speC* is also likely to explain the inability of *S.* Gallinarum to decarboxylate ornithine, a defining feature of this *Salmonella* serovar (Supplemental Table 2) (Crichton and Old 1990).

Pseudogenes potentially involved in virulence and host adaptation

S. Gallinarum 287/91 is defined as being nonmotile. *S.* Gallinarum 287/91 carries 50 genes associated with motility and chemotaxis, distributed over three loci. Of these, five genes, present in two loci, carry mutations that explain the nonmotile phenotype, including *cheM*, *flhA*, *flhB*, *flgK*, and *flgI* (Supplemental Table 1).

Of the 13 fimbrial operons detected in *S.* Enteritidis PT4, the *std* fimbrial operon is not present in *S.* Gallinarum 287/91 (see *ROD30* above). The remaining 12 *S.* Gallinarum 287/91 fimbrial operons are identical to those in *S.* Enteritidis PT4 except for the mutations in genes within operons *sti*, *stf*, *saf*, *stb*, *peg*, *lpf*, *sef*, and *sth* (see Supplemental Table 1). The level of pseudogene formation within these 12 fimbrial gene clusters (16%) is over twice that of the genome average (7%). Only operons *fim*, *bcf*, *csg*, and *ste* remain undisrupted on the *S.* Gallinarum 287/91 chromosome (summarized in Fig. 1). However, both *S.* Enteritidis and *S.* Gallinarum carry fimbrial operons on their virulence plasmids. The *S.* Enteritidis virulence plasmid carries five genes—*pefA*, *pefB*, *pefC*, *pefD*, and *pefR*—highly conserved with those of *S.* Typhimurium LT2 (Woodward and Kirwan 1996). The *pef* operon is not present on the *S.* Gallinarum 287/91 plasmid; in its place are three intact fimbrial genes displaying sequence similarity with those of the *E. coli* K88 fimbrial gene cluster (Rychlik et al. 1998). This fimbrial operon represents the only evidence of *S.* Gallinarum 287/91 having acquired new functions since the split from *S.* Enteritidis. It would be interesting to know if any isolates of *S.* Enteritidis carry such a fimbrial operon and whether this is a common characteristic of all *S.* Gallinarum strains. Significantly, the host-adapted *S.* Typhi harbors novel fimbrial genes, including Type IV pili associated with the acquisition of mobile elements (Pickard et al. 2003), and like *S.* Gallinarum 287/91, there is an elevated level of mutation in fimbrial genes (14% compared with the genome average of 4.4%), again suggesting parallel paths toward host adaptation (Table 3).

Salmonella can express several paralogous TTSS effector proteins, which show a degree of functional redundancy, for example, *sopE* and *sopE2* (Friebel et al. 2001). Other effectors related by sequence include *pipB* and *pipB2*, and *sifA* and *sifB*. *S.* Gallinarum 287/91 has lost one of each of these parlogous pairs. Other *S.* Gallinarum 287/91 TTSS effector genes that carry mutations include *sopA*, which has been implicated in *S.* Typhimurium–induced intestinal inflammation (Zhang et al. 2006). Using an antibody against the C-terminal portion of SopA, we detected a secreted protein of the expected size in *S.* Enteritidis PT4 but not *S.* Gallinarum 287/91 (data not shown), consistent with the location of a stop codon prior to the mAb-binding region in *S.* Gallinarum SopA predicted by the genome sequence (Supplemental Table 2). SopA influences *Salmonella*-induced enteritis, and taken together with the attrition of other Type III secreted effectors, this may partially dictate the differential virulence of the serovars in mammalian hosts.

As well as additional pseudogenes associated with cell interactions, again like *S.* Typhi, *S.* Gallinarum 287/91 carries mutations in genes also associated with shedding (*shdA* and *ratB*), drug resistance, DNA restriction/modification, and protective re-

Table 3. Summary of the common traits identified among the functions of genes lost independently by *S.* Typhi CT18 and *S.* Gallinarum 287/91

Process/pathway	*S.* Gallinarum 278/91[a]	*S.* Typhi CT18[b]
Cell interactions	*slrP, sopA, sifB, sspH2, sinH, pipB2, pagK, bigA*	*sopD2 (STY0971), sopA (STY2275), sopE2 (STY1987), sseJ (STY1439a), cigR (STY4024), misL (STY4030), marT (STY4027), sivH (STY2767), slrP (STY0833), bigA (STY4318)*
Fecal shedding	*shdA, ratB*	*shdA (STY2755), ratB (STY2758), sivH (STY2767)*
Fimbriae	*stdΔ, stiC, stfF, safC, stbC, pegC, lpfC, sefD, sefC, sthB, sthA, sthE*	*bcfC (STY0026), fimI (STY0590), steA (STY3084), safE (STY0333), stgC (STY3920), ushA (STY0539), sefA (STY4836a) sefD (STY4839), sefR (STY4841), sthC (STY4938), sthE (STY4938)*
Flagella/motility	*cheM, flhB, flhA, flgK, flgI*	*fliB (STY2166)*
Type I restriction modification	*hsdR, hsdM*	*hsdM (STY4833)*
Type III restriction modification		
Restriction enzyme *Sty*LT1	*mod*	*res (STY0389)*
Cobalamine biosynthesis	*pocR, cobD, cbiD, cbiC, cbiO*	*cbiM (STY2226), cbiK (STY2229), cbiJ (STY2231)*
Propanediol utilization	*pduG, pduO*	*pduN (STY2254)*
Metal/drug resistance and transport		
Copper	*cusA, cusS*	*cusA (STY0610), cusS (STY0609a)*
Nickel/cobalt	*rcnAcusA, cusS*	*rcnA (STY3169)*
Nickel	*nxiArcnA*	*nxiA (STY2901)*
Acriflavin	*acrFnxiA*	*acrE (STY3569), acrF (STY3570)*
Tetrathionate respiration	*ttrB, ttrC*	*ttrS (STY1735)*
Trehalose degradation/synthesis	*treC*	*treA (STY1924)*
Hydrogenase I	*hyaF*	*hyaB2 (STY1525), hyaA (STY1319)*
Ornithine catabolism	*speC*	*speC (STY3270), speF (STY0739)*
Amino acid catabolism		
L-serine/*L*-threonine	*tdcG*	*tdcC (STY3426)*
Cellulose biosynthesis	*bcsG*	*bcsC (STY4184)*
Surface polysaccharide		
LPS O-chain	*gtrB*	*gtrB (STY2627b)*
LPS core	*rfaZ* (*waaZ*)	*wcaK(STY2311), wcaD (STY2324), wcaA (STY2328)*
Alternative terminal electron acceptors		
Dimethyl sulfoxide reductase	*dmsA2, dmsA1*	*dmsA2 (STY4503), dmsB2 (STY4506)*
Trimethylamine *N*-oxide (TMAO)	*torS*	*torR (STY3954), torC (STY3955)*
Carbon source		
D-Glucarate uptake and degradation	*gudD*	*gudP (STY4097)*
Maltodextrin and Maltose associated	*malS, malY, malX*	*malY (STY1657a), malX (STY1657)*

[a]For systematic gene identifiers and a description of function, see Supplemental Table 1.
[b]Parkhill et al. 2001.

sponses (Supplemental Table 1). The majority of *S.* Enteritidis isolates can produce a biofilm, of which cellulose is a key component. While mutations in biofilm production may not measurably affect virulence, they are significantly less resistant to chemical and mechanical stress. Consequently, this is likely to be an adaptation by *Salmonella* to survival in the environment but has also been suggested to prolong retention in the gut (Solano et al. 2002). We have shown experimentally that *S.* Gallinarum 287/91 is unable to make cellulose, and this is likely to be explained by a mutation in *bcsG* (Supplemental Table 2). This is consistent with the reduced ability of *S.* Gallinarum to colonize the gut compared with *S.* Enteritidis.

Conclusions

The data presented in this report provide several clear messages, some of which may be experimentally tractable. Comparative analysis of the genomes of *S.* Enteritidis PT4 and *S.* Gallinarum 287/91 shows that representative strains of these two *S. enterica* serovars are highly related and that *S.* Gallinarum may be a direct descendant of *S.* Enteritidis. Importantly, *S.* Enteritidis is promiscuous, being able to colonize and infect multiple hosts, including chickens, cattle, mice, and humans, in addition to producing murine typhoid. Whereas *S.* Gallinarum is highly restricted to causing a typhoid-like disease in avian species, it is relatively noninfectious in other hosts, including mice, and does not colonize the gut of animals. Thus, we suggest that there is an experimental opportunity to use genetic approaches to define the genetic basis of host restriction by directly comparing the pathogenicity of strains of *S.* Enteritidis and *S.* Gallinarum in murine and chicken models.

Previous genome analyses on host-restricted salmonellae has involved human-restricted serovars, including *S.* Typhi and *S.* Paratyphi, limiting experimental tractability. Nevertheless genome comparisons of host-restricted/adapted *S. enterica* serovars, and indeed of other pathogens, indicate that loss of gene function may be a common evolutionary mechanism through which host adaptation occurs. Gene loss not only may limit the interhost promiscuity of the pathogen but also is likely to restrict the potential pathogenicity in the host to a more limited set of interactions. We hypothesize that gene loss may be a mechanism of targeting the invading pathogen preferentially to particular tissues or host cells and avoiding the potential stimulation of nonspecific inflammation. An example here would be the loss of flagella or fimbriae, which can mediate attachment and invasion of cell surfaces and may activate pattern recognition molecules.

In addition, gene loss can influence the ability of the pathogen to survive in the external environment or even in stressful situations within the host. Table 3 provides a list of some of the common traits identified among the functions of genes lost independently by *S.* Typhi and *S.* Gallinarum. Some of the overlaps are striking, including the loss of common TTSS effectors and genes involved in common metabolic processes such as cobalamine and propanediol utilization, tetrathionate respiration, sugar uptake and utilization, hydrogenase activity, cellulose production, ornithine decarboxylase activity, and electron transport acceptor function. Some of these common traits have also been noted to have changed in representatives of gut adapted (*Y. enterocolitica*) versus systemic (*Y. pestis)* yersiniae, and again in this system, gene loss may be involved in the adaptation from a gut to a systemic lifestyle. We believe that further studies analyzing the contribution of pseudogenes and their functional alleles to host adaptation and tissue specificity and, in particular, the parallel but overlapping degradative evolutionary pathways followed by different organisms adapting to different hosts will lead to significant understanding of the mechanisms of host adaptation and host restriction and could be applicable to the less tractable human-adapted organisms, such as *S.* Typhi.

Methods

Bacterial strains

S. Gallinarum strain 287/91 was isolated from an outbreak of fowl typhoid in brown egg–laying hens by A. Berchieri, University of Sao Paulo, Jaboticabal, Brazil. It is highly virulent (>90% mortality) in susceptible breeds of chickens (P. Barrow and A. Berchier, unpubl.). It was chosen in preference to the well-characterized strain 9 (Smith 1955) because of the length of laboratory passage of the latter strain. *S.* Enteritidis phage type 4 (PT4) strain P125109 was isolated from an outbreak of human food-poisoning in the United Kingdom that was traced back to a poultry farm. The strain is highly virulent in newly hatched chickens and is also invasive in laying hens, resulting in egg contamination (Barrow 1991; Barrow and Lovell 1991). Biochemical tests for carbohydrate catabolism were performed using api 50 CH according to manufacturer's instructions (BioMerieux).

Growth and sequencing of *S.* Enteritidis PT4 and *S.* Gallinarum 287/91

Methods for sequencing *S.* Enteritidis PT4 and *S.* Gallinarum 287/91 were identical unless stated. A single bacterial colony was picked from Congo Red agar and grown overnight in BAB broth with shaking at 37°C. Cells were collected, and total DNA (10 mg) was isolated using proteinase K treatment followed by phenol extraction. The DNA was fragmented by sonication, and several libraries were generated in pUC18 using size fractions ranging from 1.0–2.5 kb.

The whole genome sequenced to a depth of 9× coverage from M13mp18 (insert size 1.4–2 kb) and pUC18 (insert size 2.2–4.2 kb) small insert libraries using dye terminator chemistry on ABI3700 automated sequencers. End sequences from larger insert plasmid (pBACe3.6, 12–30 kb insert size) libraries were used as a scaffold.

The sequence was assembled, finished, and annotated as described previously (Parkhill et al. 2000), using the program Artemis (Berriman and Rutherford 2003) to collate data and facilitate annotation.

The genomes have been submitted to EMBL under the following accession numbers: *S.* Enteritidis PT4 genome, AM933172; *S.* Gallinarum 287/91 genome, AM933173.

In silico genome analysis

The genome sequences of *S.* Typhimurium strain LT2 (McClelland et al. 2001), *S.* Enteritidis PT4, and *S.* Gallinarum 287/91 were compared pairwise using the Artemis Comparison Tool (ACT) (Carver et al. 2005). Subsequences taken from the genomes of uropathogenic *E. coli* strain CFT073 (Welch et al. 2002), *Erwinia carotovora* sbsp. *atroseptica* strain SCRI1043 (Bell et al. 2004), and *Photorhabidus luminescens*,sbsp. *laumondii* TT01 (Duchaud et al. 2003) were compared with ACT as above and used to construct Figure 4.

Pseudogenes had one or more mutations that would ablate expression; each of the inactivating mutations was confirmed by subsequently rechecking the original sequencing data and where necessary were resequenced.

Orthologous gene sets were identified by reciprocal FASTA searches. Only those pairs of homologous CDS were retained for further analysis where the predicted amino acid identity was ≥40% over 80% of the protein length. These genes were then subject to manual curation using gene synteny to increase the accuracy of this analysis. This strategy was applied to pairwise comparisons of the genomes of *S.* Typhimurium strain LT2, *S.* Enteritidis PT4, and *S.* Gallinarum 287/91.

Cellulose production assay

For preparation and use of Calcofluor plates, Calcofluor white stain was obtained from Sigma as a 0.1% w/v solution. This was added to L-agar at a final concentration of 200 µg/mL as recommended by Solano et al. (2002). Bacterial cultures were inoculated and then left at room temperature for 48 h.

Colony fluorescence was examined by holding the plate over a 366-nm UV transilluminator. Controls used included *E. coli* C600 (negative control) and *S. typhimurium* SL1344 (positive control). Colony fluorescence was scored quantitatively using the controls as standards.

Acknowledgments

We thank the core sequencing and informatics teams at the Sanger Institute for their assistance and The Wellcome Trust for its support of the Sanger Institute Pathogen Sequencing Unit. This project was funded through The Wellcome Trust Beowulf Genomics Initiative.

References

Barrow, P. 1991. Experimental infection of chickens with *Salmonella enteritidis*. *Avian Pathol.* **20:** 145–153.

Barrow, P.A. 2000. The paratyphoid salmonellae. *Rev. Sci. Tech.* **19:** 351–375.

Barrow, P.A. and Lovell, M.A. 1991. Experimental infection of egg-laying hens with *Salmonella enteritidis*. *Avian Pathol.* **20:** 339–352.

Bausch, C., Peekhaus, N., Utz, C., Blais, T., Murray, E., Lowary, T., and Conway, T. 1998. Sequence analysis of the GntII (subsidiary) system for gluconate metabolism reveals a novel pathway for L-idonic acid catabolism in *Escherichia coli*. *J. Bacteriol.* **180:** 3704–3710.

Bell, K.S., Sebaihia, M., Pritchard, L., Holden, M.T., Hyman, L.J., Holeva, M.C., Thomson, N.R., Bentley, S.D., Churcher, L.J., Mungall, K., et al. 2004. Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors. *Proc. Natl. Acad. Sci.* **101:** 11105–11110.

Berriman, M. and Rutherford, K. 2003. Viewing and annotating sequence data with Artemis. *Brief. Bioinform.* **4:** 124–132.

Bishop, A.L., Baker, S., Jenks, S., Fookes, M., Gaora, P.O., Pickard, D., Anjum, M., Farrar, J., Hien, T.T., Ivens, A., et al. 2005. Analysis of

the hypervariable region of the *Salmonella enterica* genome associated with tRNA(*leuX*). *J. Bacteriol.* **187:** 2469–2482.
Blanc-Potard, A.B., Solomon, F., Kayser, J., and Groisman, E.A. 1999. The SPI-3 pathogenicity island of *Salmonella enterica*. *J. Bacteriol.* **181:** 998–1004.
Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G., and Parkhill, J. 2005. ACT: The Artemis Comparison Tool. *Bioinformatics* **21:** 3422–3423.
Chiu, C.H., Tang, P., Chu, C., Hu, S., Bao, Q., Yu, J., Chou, Y.Y., Wang, H.S., and Lee, Y.S. 2005. The genome sequence of *Salmonella enterica* serovar Choleraesuis, a highly invasive and resistant zoonotic pathogen. *Nucleic Acids Res.* **33:** 1690–1698.
Collighan, R.J. and Woodward, M.J. 2001. The SEF14 fimbrial antigen of *Salmonella enterica* serovar Enteritidis is encoded within a pathogenicity islet. *Vet. Microbiol.* **80:** 235–245.
Cooke, F.J., Wain, J., Fookes, M., Ivens, A., Thomson, N., Brown, D.J., Threlfall, E.J., Gunn, G., Foster, G., and Dougan, G. 2007. Prophage sequences defining hot spots of genome variation in *Salmonella enterica* serovar Typhimurium can be used to discriminate between field isolates. *J. Clin. Microbiol.* **45:** 2590–2598.
Crichton, P.B. and Old, D.C. 1990. Salmonellae of serotypes *gallinarum* and *pullorum* grouped by biotyping and fimbrial-gene probing. *J. Med. Microbiol.* **32:** 145–152.
Doyle, M., Fookes, M., Ivens, A., Mangan, M.W., Wain, J., and Dorman, C.J. 2007. An H-NS-like stealth protein aids horizontal DNA transmission in bacteria. *Science* **315:** 251–252.
Duchaud, E., Rusniok, C., Frangeul, L., Buchrieser, C., Givaudan, A., Taourit, S., Bocs, S., Boursaux-Eude, C., Chandler, M., Charles, J.F., et al. 2003. The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. *Nat. Biotechnol.* **21:** 1307–1313.
Edwards, R.A., Schifferli, D.M., and Maloy, S.R. 2000. A role for *Salmonella* fimbriae in intraperitoneal infections. *Proc. Natl. Acad. Sci.* **97:** 1258–1262.
Eriksson, S., Lucchini, S., Thompson, A., Rhen, M., and Hinton, J.C. 2003. Unravelling the biology of macrophage infection by gene expression profiling of intracellular *Salmonella enterica*. *Mol. Microbiol.* **47:** 103–118.
Figueroa-Bossi, N., Uzzau, S., Maloriol, D., and Bossi, L. 2001. Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in *Salmonella*. *Mol. Microbiol.* **39:** 260–271.
Friebel, A., Ilchmann, H., Aepfelbacher, M., Ehrbar, K., Machleidt, W., and Hardt, W.D. 2001. SopE and SopE2 from *Salmonella typhimurium* activate different sets of RhoGTPases of the host cell. *J. Biol. Chem.* **276:** 34035–34040.
Hansen-Wester, I. and Hensel, M. 2001. Salmonella pathogenicity islands encoding type III secretion systems. *Microbes Infect.* **3:** 549–559.
Hensel, M. 2000. *Salmonella* pathogenicity island 2. *Mol. Microbiol.* **36:** 1015–1023.
Iqbal, M., Philbin, V.J., Withanage, G.S., Wigley, P., Beal, R.K., Goodchild, M.J., Barrow, P., McConnell, I., Maskell, D.J., Young, J., et al. 2005. Identification and functional characterization of chicken toll-like receptor 5 reveals a fundamental role in the biology of infection with *Salmonella enterica* serovar typhimurium. *Infect. Immun.* **73:** 2344–2350.
Kaiser, P., Rothwell, L., Galyov, E.E., Barrow, P.A., Burnside, J., and Wigley, P. 2000. Differential cytokine expression in avian cells in response to invasion by *Salmonella typhimurium*, *Salmonella enteritidis* and *Salmonella gallinarum*. *Microbiology* **146:** 3217–3226.
Kilger, G. and Grimont, P.A. 1993. Differentiation of *Salmonella* phase 1 flagellar antigen types by restriction of the amplified *fliC* gene. *J. Clin. Microbiol.* **31:** 1108–1110.
Klumpp, J. and Fuchs, T.M. 2007. Identification of novel genes in genomic islands that contribute to *Salmonella typhimurium* replication in macrophages. *Microbiology* **153:** 1207–1220.
Li, J., Smith, N.H., Nelson, K., Crichton, P.B., Old, D.C., Whittam, T.S., and Selander, R.K. 1993. Evolutionary origin and radiation of the avian-adapted non-motile salmonellae. *J. Med. Microbiol.* **38:** 129–139.
McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., et al. 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* **413:** 852–856.
McClelland, M., Sanderson, K.E., Clifton, S.W., Latreille, P., Porwollik, S., Sabo, A., Meyer, R., Bieri, T., Ozersky, P., McLellan, 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:** 1268–1274.
McMeechan, A., Lovell, M.A., Cogan, T.A., Marston, K.L., Humphrey, T.J., and Barrow, P.A. 2005. Glycogen production by different *Salmonella enterica* serotypes: Contribution of functional *glgC* to virulence, intestinal colonization and environmental survival. *Microbiology* **151:** 3969–3977.
Mmolawa, P.T., Schmieger, H., Tucker, C.P., and Heuzenroeder, M.W. 2003. Genomic structure of the *Salmonella enterica* serovar Typhimurium DT 64 bacteriophage ST64T: Evidence for modular genetic architecture. *J. Bacteriol.* **185:** 3473–3475.
Navarre, W.W., Porwollik, S., Wang, Y., McClelland, M., Rosen, H., Libby, S.J., and Fang, F.C. 2006. Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science* **313:** 236–238.
Nishino, K. and Yamaguchi, A. 2001. Analysis of a complete library of putative drug transporter genes in *Escherichia coli*. *J. Bacteriol.* **183:** 5803–5812.
Parkhill, J., Wren, B.W., Mungall, K., Ketley, J.M., Churcher, C., Basham, D., Chillingworth, T., Davies, R.M., Feltwell, T., Holroyd, S., et al. 2000. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* **403:** 665–668.
Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Wain, J., Churcher, C., Mungall, K.L., Bentley, S.D., Holden, M.T., et al. 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* **413:** 848–852.
Pickard, D., Wain, J., Baker, S., Line, A., Chohan, S., Fookes, M., Barron, A., Gaora, P.O., Chabalgoity, J.A., Thanky, N., et al. 2003. Composition, acquisition, and distribution of the Vi exopolysaccharide-encoding *Salmonella enterica* pathogenicity island SPI-7. *J. Bacteriol.* **185:** 5055–5065.
Raleigh, E.A. 1992. Organization and function of the *mcrBC* genes of *Escherichia coli* K-12. *Mol. Microbiol.* **6:** 1079–1086.
Reitzer, L. 2003. Nitrogen assimilation and global regulation in *Escherichia coli*. *Annu. Rev. Microbiol.* **57:** 155–176.
Rodrigue, D.C., Tauxe, R.V., and Rowe, B. 1990. International increase in *Salmonella enteritidis*: A new pandemic? *Epidemiol. Infect.* **105:** 21–27.
Roth, J.R., Lawrence, J.G., and Bobik, T.A. 1996. Cobalamin (coenzyme B12): synthesis and biological significance. *Annu. Rev. Microbiol.* **50:** 137–181.
Rychlik, I., Lovell, M.A., and Barrow, P.A. 1998. The presence of genes homologous to the K88 genes *faeH* and *faeI* on the virulence plasmid of *Salmonella gallinarum*. *FEMS Microbiol. Lett.* **159:** 255–260.
Schneider, E., Freundlieb, S., Tapio, S., and Boos, W. 1992. Molecular characterization of the MalT-dependent periplasmic alpha-amylase of *Escherichia coli* encoded by *malS*. *J. Biol. Chem.* **267:** 5148–5154.
Shah, D.H., Lee, M.J., Park, J.H., Lee, J.H., Eo, S.K., Kwon, J.T., and Chae, J.S. 2005. Identification of *Salmonella gallinarum* virulence genes in a chicken infection model using PCR-based signature-tagged mutagenesis. *Microbiology* **151:** 3957–3968.
Shivaprasad, H.L. 2000. Fowl typhoid and pullorum disease. *Rev. Sci. Tech.* **19:** 405–424.
Smith, H.W. 1955. Observations on experimental fowl typhoid. *J. Comp. Pathol.* **65:** 37–54.
Solano, C., Garcia, B., Valle, J., Berasain, C., Ghigo, J.M., Gamazo, C., and Lasa, I. 2002. Genetic analysis of *Salmonella enteritidis* biofilm formation: Critical role of cellulose. *Mol. Microbiol.* **43:** 793–808.
Stanley, T.L., Ellermeier, C.D., and Slauch, J.M. 2000. Tissue-specific gene expression identifies a gene in the lysogenic phage Gifsy-1 that affects *Salmonella enterica* serovar Typhimurium survival in Peyer's patches. *J. Bacteriol.* **182:** 4406–4413.
Sweeney, N.J., Laux, D.C., and Cohen, P.S. 1996. *Escherichia coli* F-18 and *E. coli* K-12 *eda* mutants do not colonize the streptomycin-treated mouse large intestine. *Infect. Immun.* **64:** 3504–3511.
Thomson, N., Baker, S., Pickard, D., Fookes, M., Anjum, M., Hamlin, N., Wain, J., House, D., Bhutta, Z., Chan, K., et al. 2004. The role of prophage-like elements in the diversity of *Salmonella enterica* serovars. *J. Mol. Biol.* **339:** 279–300.
Thomson, N.R., Howard, S., Wren, B.W., Holden, M.T., Crossman, L., Challis, G.L., Churcher, C., Mungall, K., Brooks, K., Chillingworth, T., et al. 2006. The complete genome sequence and comparative genome analysis of the high pathogenicity *Yersinia enterocolitica* strain 8081. *PLoS Genet.* **2:** e206. doi: 10.1371/journal.pgen.0020206.
Titheradge, A.J., Ternent, D., and Murray, N.E. 1996. A third family of allelic hsd genes in *Salmonella enterica*: Sequence comparisons with related proteins identify conserved regions implicated in restriction of DNA. *Mol. Microbiol.* **22:** 437–447.
Townsend, S.M., Kramer, N.E., Edwards, R., Baker, S., Hamlin, N., Simmonds, M., Stevens, K., Maloy, S., Parkhill, J., Dougan, G., et al. 2001. *Salmonella enterica* serovar Typhi possesses a unique repertoire of fimbrial gene sequences. *Infect. Immun.* **69:** 2894–2901.
Vernikos, G.S. and Parkhill, J. 2006. Interpolated variable order motifs for identification of horizontally acquired DNA: Revisiting the

Salmonella pathogenicity islands. *Bioinformatics* **22:** 2196–2203.
Welch, R.A., Burland, V., Plunkett III, G., Redford, P., Roesch, P., Rasko, D., Buckles, E.L., Liou, S.R., Boutin, A., Hackett, J., et al. 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci.* **99:** 17020–17024.
Williamson, H.S. and Free, A. 2005. A truncated H-NS-like protein from enteropathogenic *Escherichia coli* acts as an H-NS antagonist. *Mol. Microbiol.* **55:** 808–827.
Wong, K.K., McClelland, M., Stillwell, L.C., Sisk, E.C., Thurston, S.J., and Saffer, J.D. 1998. Identification and sequence analysis of a 27-kilobase chromosomal fragment containing a *Salmonella* pathogenicity island located at 92 minutes on the chromosome map of *Salmonella enterica* serovar typhimurium LT2. *Infect. Immun.* **66:** 3365–3371.
Wood, M.W., Jones, M.A., Watson, P.R., Hedges, S., Wallis, T.S., and Galyov, E.E. 1998. Identification of a pathogenicity island required for *Salmonella* enteropathogenicity. *Mol. Microbiol.* **29:** 883–891.
Woodward, M.J. and Kirwan, S.E. 1996. Detection of *Salmonella enteritidis* in eggs by the polymerase chain reaction. *Vet. Rec.* **138:** 411–413.
Zhang, Y., Higashide, W.M., McCormick, B.A., Chen, J., and Zhou, D. 2006. The inflammation-associated *Salmonella* SopA is a HECT-like E3 ubiquitin ligase. *Mol. Microbiol.* **62:** 786–793.

Received February 12, 2008; accepted in revised form June 17, 2008.

Comparative genome analysis of *Salmonella* Enteritidis PT4 and *Salmonella* Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways

Nicholas R. Thomson, Debra J. Clayton, Daniel Windhorst, et al.

Genome Res. 2008 18: 1624-1637 originally published online June 26, 2008
Access the most recent version at doi:10.1101/gr.077404.108