

27. Bar, M. & Biederman, I. Subliminal visual priming. *Psychol. Sci.* **9**, 464–469 (1998).  
 28. Chun, M. M. Contextual cueing of visual attention. *Trends Cogn. Sci.* **4**, 170–178 (2000).  
 29. Barlow, H. B. in *Vision: Coding and Efficiency* (ed. Blakemore, C.) (Cambridge Univ. Press, Cambridge, 1989).

## Acknowledgements

We thank M. Ahissar, P. Cavanagh, S. Hochstein, W. T. Newsome, R. Raizada and R. Sekuler for their comments on this or related study. This work was supported by an NSF (SBES, HCP) grant to T.W. and by Chiba University and JSPS fellowship to Y.S.

Correspondence and requests for materials should be addressed to T.W.  
 (e-mail: takeo@bu.edu).

## Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18

J. Parkhill\*, G. Dougan†, K. D. James\*, N. R. Thomson\*, D. Pickard†, J. Wain†, C. Churcher\*, K. L. Mungall\*, S. D. Bentley\*, M. T. G. Holden\*, M. Sebaihia\*, S. Baker\*, D. Basham\*, K. Brooks\*, T. Chillingworth\*, P. Connerton†, A. Cronin\*, P. Davis\*, R. M. Davies\*, L. Dowd\*, N. White‡, J. Farrar‡, T. Feltwell\*, N. Hamlin\*, A. Haque†, T. T. Hien§, S. Holroyd\*, K. Jagels\*, A. Krogh||, T. S. Larsen||, S. Leather\*, S. Moule\*, P. Ó'Gaora†, C. Parry§, M. Quail\*, K. Rutherford\*, M. Simmonds\*, J. Skelton\*, K. Stevens\*, S. Whitehead\* & B. G. Barrell\*

\* The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

† Centre for Molecular Microbiology and Infection, Department of Biological Sciences, Imperial College of Science, Technology and Medicine, London SW7 2AZ, UK

‡ University of Oxford-Wellcome Trust Clinical Research Unit, Centre for Tropical Diseases, 190 Ben Ham Thu, Quan 5, Ho Chi Minh City, Vietnam, and Centre for Tropical Medicine, University of Oxford, OX3 9DU, UK

§ The Centre for Tropical Diseases, 190 Ben Ham Thu, Quan 5, Ho Chi Minh City, Vietnam

|| Centre for Biological Sequence Analysis, BioCentrum-DTU, The Technical University of Denmark, Building 208, 2800 Lyngby, Denmark

*Salmonella enterica* serovar Typhi (*S. typhi*) is the aetiological agent of typhoid fever, a serious invasive bacterial disease of humans with an annual global burden of approximately 16 million cases, leading to 600,000 fatalities<sup>1</sup>. Many *S. enterica* serovars actively invade the mucosal surface of the intestine but are normally contained in healthy individuals by the local immune defence mechanisms. However, *S. typhi* has evolved the ability to spread to the deeper tissues of humans, including liver, spleen and bone marrow. Here we have sequenced the 4,809,037-base pair (bp) genome of a *S. typhi* (CT18) that is resistant to multiple drugs, revealing the presence of hundreds of insertions and deletions compared with the *Escherichia coli* genome, ranging in size from single genes to large islands. Notably, the genome sequence identifies over two hundred pseudogenes, several corresponding to genes that are known to contribute to virulence in *Salmonella typhimurium*. This genetic degradation may contribute to the human-restricted host range for *S. typhi*. CT18 harbours a 218,150-bp multiple-drug-resistance *incH1* plasmid (pHCM1), and a 106,516-bp cryptic plasmid (pHCM2), which shows recent common ancestry with a virulence plasmid of *Yersinia pestis*.

*Salmonella typhi* is a serovar of *S. enterica*, which is serologically O (lipopolysaccharide) type 09, 012; H (flagellin) type d; and Vi

(extracellular capsule) positive. Humans are the only known natural host of *S. typhi*, with *S. typhi* showing limited pathogenicity for most animals. Isoenzyme analysis has suggested that isolates of *S. typhi* around the world are highly related<sup>2</sup>, a view confirmed using multi-locus sequence analysis (C. Kidgell *et al.*, unpublished data). Multiple drug resistance (MDR) is a serious emerging threat to the treatment of infectious diseases—MDR *S. typhi* are resistant to commonly available antibiotics, and clinical resistance to fluoroquinolones, the most effective antimicrobials for the treatment of typhoid fever, has been reported<sup>3</sup>. *Salmonella typhi* CT18 is an example of an emerging MDR microorganism; in depth genome analysis will contribute to our understanding of how such microorganisms adapt rapidly to new environmental opportunities that are presented by modern human society.

The principal features of the *S. typhi* CT18 chromosome and the two plasmids harboured by this strain are shown in Table 1. The beginning of the sequence was taken to correspond with minute 0 on the *E. coli* and *Salmonella* genetic maps<sup>4</sup>; the origin and terminus of replication, predicted by comparison with *E. coli* and confirmed by GC-bias (Fig. 1), are near 3.765 megabases (Mb) and 1.437 Mb, respectively. The metabolism of *Salmonella* and *E. coli* has been extensively studied over many decades<sup>4</sup>, and our analysis reveals few surprises in this area. However, the chromosome was predicted to encode 204 pseudogenes, which is remarkable in a genome of an organism capable of growth both inside and outside the host. Most of these pseudogenes (124 out of 204) have been inactivated by the introduction of a single frameshift or stop codon, which suggests that they are of recent origin. Five pseudogenes (*priC*, *ushA*, *fepE*, *sopE2* and *fliB*) were re-sequenced from several independent *S. typhi* isolates, and were identical in every case. Frameshifts that are due to changes in the length of homopolymeric tracts account for 45 pseudogenes; this is a mechanism of variation that was previously shown to occur in *E. coli*<sup>5</sup>, although at much lower rates than are required for rapid phase variation in other organisms. Some of the pseudogenes (27 out of 204) are the remnants of insertion sequence (IS) transposases, integrases and genes of bacteriophage origin. However, there are a significant number (75 out of 204) that are predicted to be involved in housekeeping functions, such as a component of the DNA primase complex, *priC*, cobalamin biosynthesis genes *chiM*, *-J*, *-K* and *-C*, the proline transporter *proV*, and the anaerobic dimethyl sulphoxide reductase components *dmsA* and *dmsB*. Many more mutations (46 out of 204) are in genes that are potentially involved in virulence or host interaction. Examples of this latter group include components of seven of the twelve

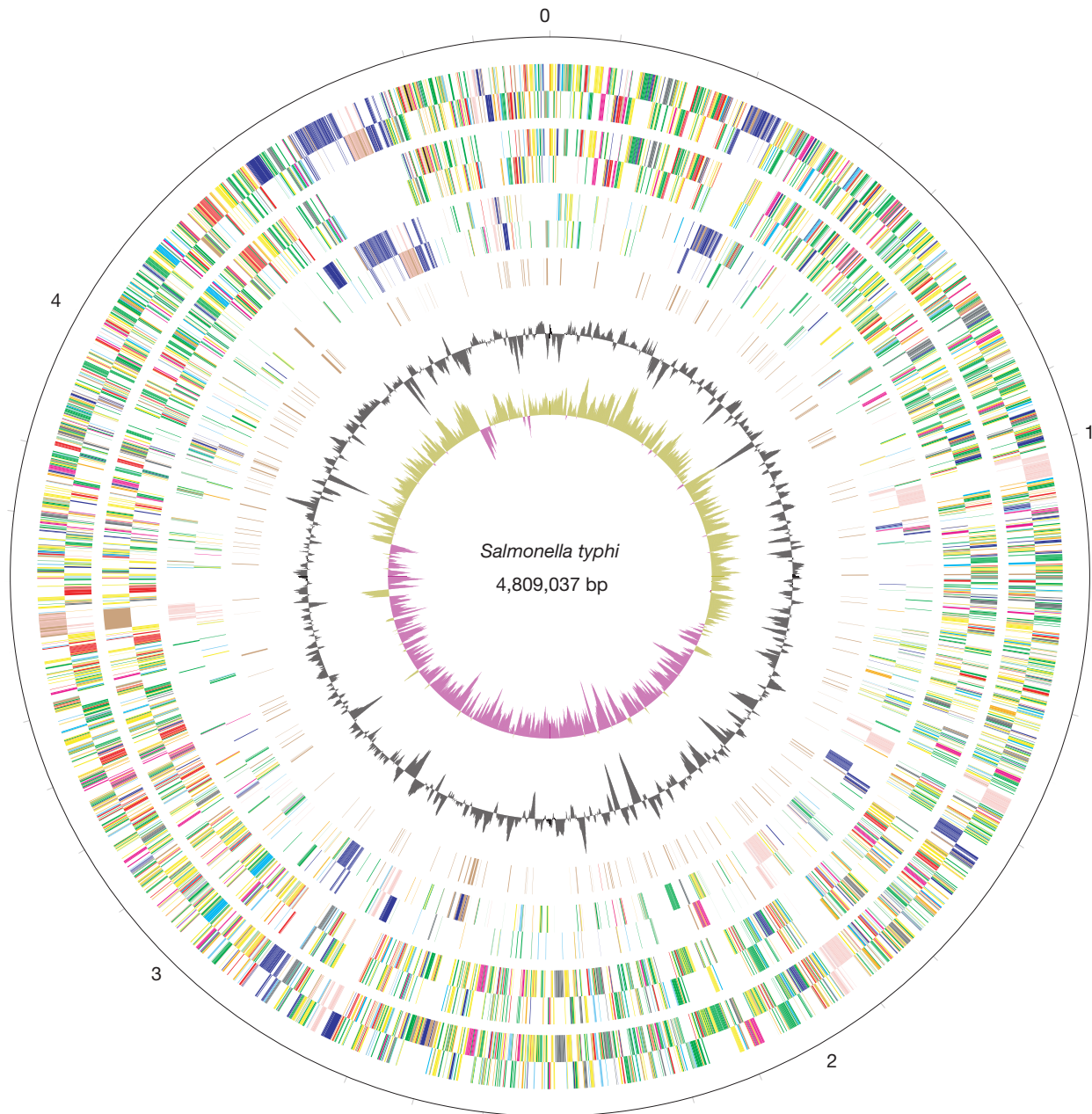
**Table 1 Features of the *S. typhi* genome**

Component of genome	Property
Chromosome	
Total size	4,809,037 bp
G+C content	52.09%
Coding sequences	4,599
... of which pseudogenes	204
Coding density	87.6%
Average gene length	958 bp
Ribosomal RNAs	6 × (16S–23S–5S), 1 × (16S–23S–5S–5S)
Transfer RNAs	78
Other stable RNAs	8
pHCM1	
Total size	218,150 bp
G+C content	47.58%
Coding sequences	249
... of which pseudogenes	8
Coding density	83.8%
Average gene length	759 bp
pHCM2	
Total size	106,516 bp
G+C content	50.6%
Coding sequences	131
... of which pseudogenes	0
Coding density	87.1%
Average gene length	708 bp
Transfer RNAs	1

chaperone–usher fimbrial operons<sup>6</sup>; the gene responsible for flagellar methylation, *fliB*; genes within or associated with previously described *Salmonella* pathogenicity islands (SPI) (for example, the sensor kinase *ttrS* that is associated with SPI-2 (ref. 7) and *cigR*, *marT* and *misL* from SPI-3 (ref. 8); the leucine-rich repeat protein *slrP*, which is involved in host-range specificity<sup>9</sup> and is secreted through a type III system; other type-III-secreted effector proteins including *sseJ* (ref. 10), *sopE2* (ref. 11) and *sopA*; and the genes *shdA*, *ratA* and *sivH*, which are present in an island unique to *Salmonellae* infecting warm-blooded vertebrates<sup>12</sup>. A greater proportion of pseudogenes than expected lie within islands that are unique to *S. typhi* compared with *E. coli* (59% (120 out of 204) compared with 33% (1,505 out of 4,599) of all genes). This apparent inactivation of

many of the mechanisms of host interaction may go some way to explaining the host restriction of *S. typhi* relative to other *Salmonella* serovars, and suggests that *S. typhi* may have passed through a recent evolutionary bottleneck.

Apart from some previously detected rearrangements<sup>13</sup>, which were caused by recombination between ribosomal RNA operons near the origin of replication, the genomes of *S. typhi* and *E. coli*<sup>14</sup> are essentially collinear along their entire length (see Supplementary Information for a linear gene map and functional classification of identified genes; see also [http://www.sanger.ac.uk/Projects/S\\_typhi](http://www.sanger.ac.uk/Projects/S_typhi)). Although there are some cases of translocation of small gene blocks, most of the differences are due to insertions, deletions or replacements. Several of the large insertions that are present in



**Figure 1** Circular representation of the *S. typhi* genome. The outer scale is marked in megabases. Circles range from 1 (outer circle) to 9 (inner circle). Circles 1 and 2, genes on forward and reverse strand; circles 3 and 4, genes conserved with *E. coli*; circles 5 and 6, genes unique to *S. typhi* with respect to *E. coli*; circle 7, pseudogenes; circle 8, G+C content; circle 9, GC bias ((G–C/G+C); khaki indicates values >1; purple <1). All

genes are colour-coded by function: dark blue, pathogenicity/adaptation; black, energy metabolism; red, information transfer; dark green, membranes/surface structures; cyan, degradation of macromolecules; purple, degradation of small molecules; yellow, central/intermediary metabolism; light blue, regulators; pink, phage/IS elements; orange, conserved hypothetical; pale green, unknown function; brown, pseudogenes.

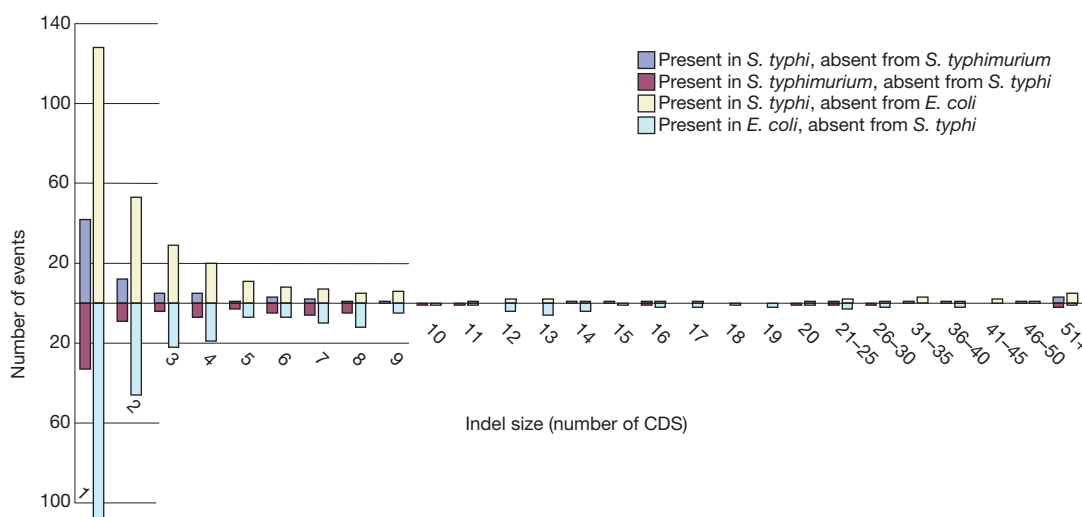
*Salmonella* have been extensively studied. These insertions generally carry genes that are important for survival in the host (including two type III secretion systems and an array of effector proteins and metal-ion transporters), they have often been inserted adjacent to a stable RNA gene, and may carry a gene encoding an integrase or transposase-like protein<sup>15</sup>. These regions, termed SPIs, are believed to be fairly recent horizontal acquisitions, and may be self-mobile. However, as observed in the comparison between *E. coli* K12 and 0157:H7 (ref. 16), there are many more insertions and deletions, including much smaller blocks (Fig. 2). Together there are 1,505 genes (32.7%) in 290 blocks that are unique to *S. typhi* relative to *E. coli* (see Supplementary Information), and 1,220 genes (28.4%) in 268 blocks that are unique to *E. coli* relative to *S. typhi*. Single-gene insertions account for 128 of the unique genes in *S. typhi*, and 456 genes are in insertions of 5 genes or less.

Among the larger blocks are the previously characterized SPIs 1–5 and seven prophage elements. There are also at least five more islands that have the characteristics of SPIs. Of the previously published SPIs we found that the original sequence of SPI-4 (ref. 17) varied most markedly from our sequence. SPI-4 was originally predicted to encode 18 genes, but our analysis of this region revealed the presence of only 8 coding sequences (CDS), two of which (STY4458 and STY4459) spanned most of those originally observed (these are represented by just one CDS in *S. typhimurium*). SPI-4 carries three CDS that are predicted to encode a type I secretion system—STY4458 and STY4459 are large, highly repetitive, and are weakly similar to RTX-like toxins. Of the new SPIs, SPI-6 (59 kb) encodes the *safA-D* and *tcsA-R* chaperone–usher fimbrial operons<sup>6</sup>. SPI-7 (134 kb) encodes the Vi biosynthetic genes<sup>18</sup>, the *SopE* prophage<sup>19</sup> and a type IVB pilus operon<sup>20</sup>. SPI-8 (6.8 kb) encodes two bacteriocin pseudogenes (STY3280 and STY3282) and a degenerate integrase; notably, genes conferring immunity to the bacteriocins remain intact. SPI-9 (16 kb), like SPI-4, encodes a type I secretory apparatus and a single, large RTX-like protein (STY2875). SPI-10 (33 kb) carries phage 46 and the *sefA-R* chaperone–usher fimbrial operon. In addition to these large islands, there are many insertions of smaller gene blocks and individual genes that may be involved in pathogenicity. These include: numerous secreted and integral membrane proteins; many new regulators; a set of eight genes that are potentially involved in extracellular polysaccharide biosynthesis (STY0759–0768, three of which are pseudogenes); a predicted iron-uptake ABC transporter (STY0802–0803); two puta-

tive efflux pumps (STY0278 and STY0414); and genes encoding several secreted effector proteins, including SifA, SopD, SopD2 and SspH2.

As would be expected, the relationship between *S. typhi* and *S. typhimurium*<sup>21</sup> is very much closer than that between *S. typhi* and *E. coli*, although there are still significant differences. The same conservation of gene order is apparent (see Supplementary Information), disrupted only by the rearrangements around the rRNAs and a large inversion around the terminus in *S. typhimurium*<sup>13</sup>. As with *E. coli*, the differences are not limited to a few large blocks. Single-gene insertions account for 42 unique genes in *S. typhi*, and 103 genes are in insertions of 5 genes or less. In all, there are 601 genes (13.1%) in 82 blocks that are unique to *S. typhi* compared with *S. typhimurium* (see Supplementary Information), and 479 genes (10.9%) in 80 blocks that are unique to *S. typhimurium* relative to *S. typhi*. Several significant insertions are apparent in *S. typhi*. These insertions include: the *staA-G*, *tcfA-D*, *steA-G* and *stgA-G* chaperone–usher fimbrial systems<sup>6</sup>; a homologue of the *E. coli* haemolysin E (STY1498); a short insert carrying homologues of the *Campylobacter* toxin *cdtB* (STY1886) and the *Bordetella pertussis* toxin genes *ptxA* and *ptxB* (STY1890 and STY1891); a putative polysaccharide acetyltransferase (STY2629); phages 10, 15, 18 and 46; and SPIs 7, 8 and 10. Aside from these relatively few insertions, much of the difference in phenotype and host range between *S. typhi* and *S. typhimurium* may well be explained by the pseudogenes described above. Of the 204 pseudogenes that have been discovered in *S. typhi*, 145 are present as intact genes in *S. typhimurium*, whereas only 23 are present as pseudogenes (15 of which have the same inactivating mutations).

*Salmonella typhi* CT18 harbours two plasmids, the larger of which (pHCM1) is conjugative and encodes resistances to all of the first-line drugs used for the treatment of typhoid fever. pHCM1 shares around 168 kb of DNA at greater than 99% sequence identity with R27 (ref. 22)—an *incH1* plasmid that was first isolated in the 1960s from *S. enterica*. Shared functions include plasmid transfer and partition, and pHCM1 has apparently been derived by the insertion of 46 CDS, totalling around 50 kb, primarily at two positions in an R27-like ancestral plasmid (Fig. 3). Eighteen of these CDS are involved with resistance to antimicrobial agents or heavy metals, and 16 are of unknown function—just three of which are similar to other plasmid-borne genes. Several intact and degenerate integrases and transposases are clustered around these two regions of insertion,



**Figure 2** Distribution of insertions and deletions in *S. typhi* relative to *E. coli* and *S. typhimurium*. The graph shows number of insertion/deletion events plotted against the size of the inserted or deleted element (shown as number of genes), clearly indicating that

most of the events involve a small number of genes. Values above the lines represent genes present in *S. typhi*; values below the line represent genes absent in *S. typhi*. Dark bars show the comparison with *S. typhimurium*; light bars with *E. coli*.

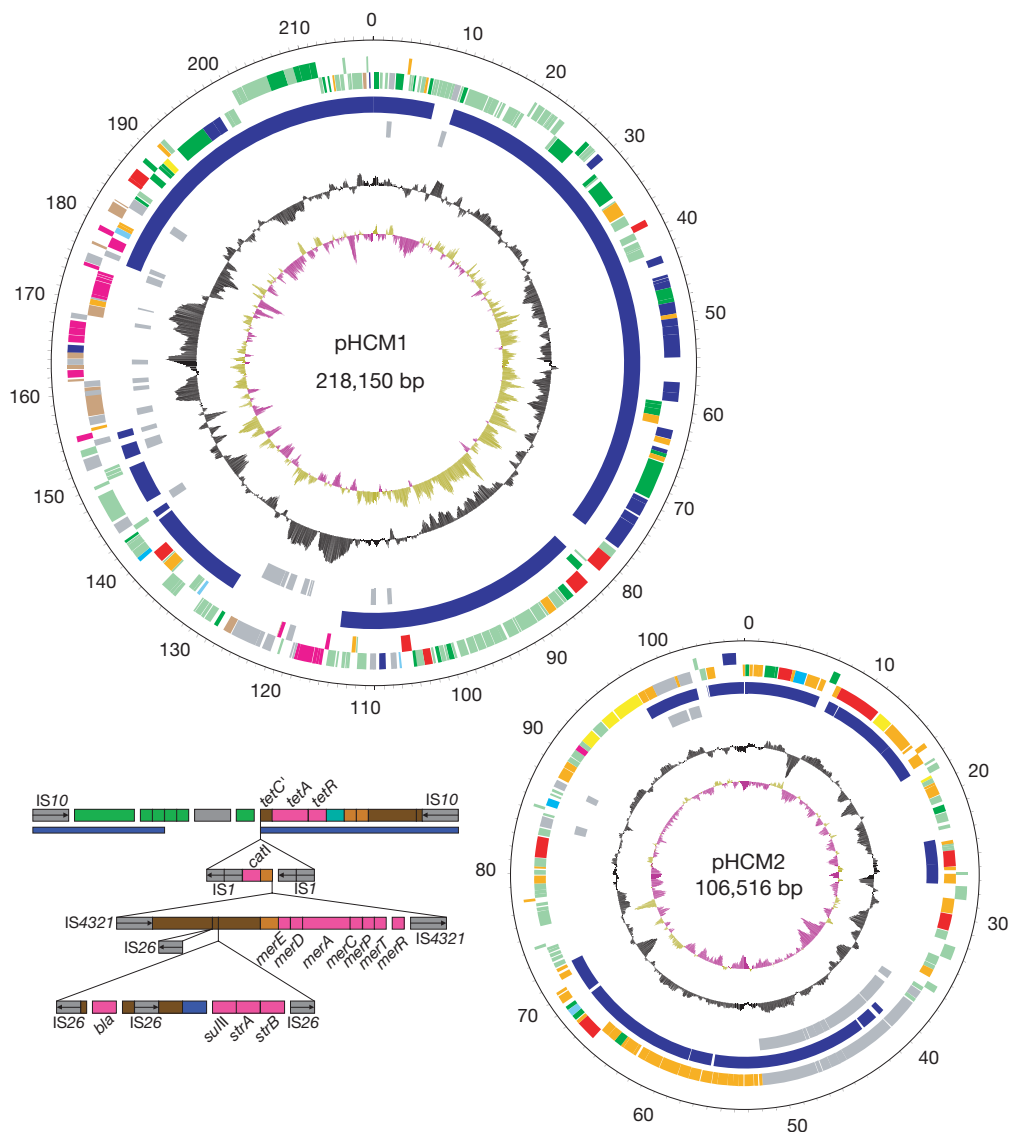


and mercury resistance operons have been acquired independently in each. Many clinically relevant resistance genes were identified, including *dhfr1b* (trimethoprim), *sulII* (sulphonamide), *catI* (chloramphenicol), *bla* (TEM-1; ampicillin) and *strAB* (streptomycin). The last four resistance determinants appear to have been inserted into the *tetC* gene of the R27 tetracycline resistance operon in several sequential IS element-mediated events (Fig. 3c). Although it has been reported that MDR plasmids bestow enhanced clinical virulence<sup>23</sup>, we can find no obvious virulence-associated genes on pHCM1.

pHCM2 is phenotypically cryptic, yet it shares over 56% of its sequence (at greater than 97% DNA identity) with the *Yersinia pestis* virulence-associated plasmid pMT1 (ref. 24). pMT1 encodes major virulence-associated determinants of *Y. pestis*, and the acquisition of this plasmid was a significant step in the evolution of the plague bacilli<sup>25</sup>. However, pHCM2 lacks the capsular antigen operon (*cafI*) and murine toxin genes that are characteristic of *Y. pestis* pMT1. The sequences that are shared between pMT1 and pHCM1 are not

contiguous but fall into several blocks (Fig. 3). Examination of the G+C content of the unique and conserved sequences of pMT1 and pHCM2 suggests that pMT1 may have been derived from a pHCM2-like precursor plasmid<sup>26</sup>. We have detected plasmids related to pHCM2 in *S. typhi* only from Southeast Asia, but most *S. typhi* do not harbour this plasmid (data not shown). The CDS that are unique to pHCM2 show similarities to a number of bacteriophage genes and to genes directly or indirectly involved in DNA biosynthesis and replication. These include a gene cluster that encodes genes similar to thymidylate synthetase, dihydrofolate reductase, ribonuclease H and ribonucleotide diphosphate reductase, as well as a putative primosomal gene cluster. In bacteriophage T4 these genes form an integral part of the primase replication complex<sup>27</sup> that facilitates rapid phage DNA biosynthesis and replication.

The sequence of the *S. typhi* genome, together with *E. coli* K12 (ref. 14) and 0157:H7 (ref. 16), reveals an unexpectedly large diversity in gene complements among these organisms. Much of



**Figure 3** Circular representations of pHCM1 and pHCM2. For the circular diagrams, the outer scale is marked in kilobases. Circles are numbered to the same scheme as in Fig. 1. Circles 1 and 2, all genes; circle 3, regions conserved in R27 (for pHCM1) or pMT1 (for pHCM2); circle 4, phage and IS genes; circle 5, G+C content; circle 6, GC bias ((G – C/G + C); khaki indicates values >1; purple <1). The linear figure represents one possible order of IS-mediated acquisition of resistance determinants in pHCM1 on the

basis of likely pairs of IS elements and interrupted genes. Colour-coding for genes: pale green, unknown function; orange, conserved hypothetical; grey, bacteriophage/IS elements; brown, pseudogenes; red, DNA metabolism; dark green, membrane/surface structures; yellow, metabolic genes; cyan, nucleases; purple, resistance determinants; dark blue, plasmid functions.

this diversity is located in discrete gene clusters that are spread throughout the different genomes. In contrast to this diversity, these enteric microorganisms exhibit marked synteny in their large-scale genomic organization, bearing in mind that *E. coli* and *S. enterica* diverged about 100 Myr ago<sup>28</sup>. The conserved genes may be a reflection of the basic lifestyle of the bacteria, requiring intestine colonization, environmental survival and transmission. The unique gene clusters probably contribute to adaptation to environmental niches and to pathogenicity. The pseudogene complement of *S. typhi* has implications for our understanding of the tight host restriction of this organism, and raises the question of whether it may be possible to eradicate *S. typhi* and typhoid fever altogether. □

## Methods

*Salmonella typhi* CT18 was isolated in December 1993, at the Mekong Delta region of Vietnam, from a 9-year-old girl who was suffering from typhoid. The strain was isolated from blood using routine culture methods<sup>23</sup>, and after serological and metabolic confirmation of the strain as *S. typhi* it was immediately frozen in glycerol at -70 °C. The genome sequence was obtained from 97,000 end sequences (giving 7.9× coverage) derived from several pUC18 genomic shotgun libraries (with insert sizes ranging from 1.4 to 4.0 kb) using dye terminator chemistry on ABI377 automated sequencers. This was supplemented with 0.7× sequence coverage from M13mp18 libraries with similar insert sizes. End sequences from a larger insert plasmid (pSP64; 1.9× clone coverage, 10–14-kb insert size) and lambda (lambda-FIX-II; 0.4× clone coverage, 20–22-kb insert size) libraries were used as a scaffold, and the final assembly was verified by comparison with restriction-enzyme digest patterns using pulsed-field gel electrophoresis (data not shown). Total sequence coverage was 9.1×. The sequence was assembled, finished and annotated as described<sup>29</sup>, using Artemis<sup>30</sup> to collate data and facilitate annotation. In addition we used a genefinder that was trained specifically for *S. typhi*, which uses a hidden Markov model with modules for the coding region, start and stop codons, and the ribosome-binding site (T.S.L. and A.K., unpublished data). The genome and proteome sequences of *S. typhi* and *S. typhimurium* or *E. coli* were compared in parallel to identify deletions and insertions using the Artemis Comparison Tool (ACT) (K. Rutherford, unpublished data; see also <http://www.sanger.ac.uk/Software/ACT/>). Pseudogenes had one or more mutations that would ablate expression, and were identified by direct comparison with *S. typhimurium*; each of the inactivating mutations was subsequently checked against the original sequencing data.

Received 4 April; accepted 20 July 2001.

- Ivanhoff, B. Typhoid fever, global situation and WHO recommendations. *Southeast Asian J. Trop. Med. Public Health* **26**, (Suppl. 2) 1–6 (1995).
- Reeves, M. W., Evins, G. M., Heiba, A. A., Plikaytis, B. D. & Farmer, J. J. Clonal nature of *Salmonella typhi* and its genetic relatedness to other salmonellae as shown by multilocus enzyme electrophoresis, and proposal of *Salmonella bongori* comb. nov. *J. Clin. Microbiol.* **27**, 313–230 (1989).
- Parry, C., Wain, J., Chinh, N. T., Vinh, H. & Farrar, J. J. Quinolone-resistant *Salmonella typhi* in Vietnam. *Lancet* **351**, 1289 (1998).
- Neidhardt, F. C. & Curtiss, R. *Escherichia coli* and *Salmonella: Cellular and Molecular Biology* (ASM Press, Washington DC, 1996).
- Rosenberg, S. M., Longrich, S., Gee, P. & Harris, R. S. Adaptive mutation by deletions in small mononucleotide repeats. *Science* **265**, 405–407 (1994).
- Townsend, S. M. *et al.* *Salmonella enterica* serovar Typhi possesses a unique repertoire of fimbrial gene sequences. *Infect. Immun.* **69**, 2894–2901 (2001).
- Hensel, M., Hensley, A. P., Nikolaus, T., Sawers, G. & Berks, B. C. The genetic basis of tetrathionate respiration in *Salmonella typhimurium*. *Mol. Microbiol.* **32**, 275–287 (1999).
- Blanc-Potard, A. B., Solomon, F., Kayser, J. & Groisman, E. A. The SPI-3 pathogenicity island of *Salmonella enterica*. *J. Bacteriol.* **181**, 998–1004 (1999).
- Tsolis, R. M. *et al.* Identification of a putative *Salmonella enterica* serotype typhimurium host range factor with homology to IpaH and YopM by signature-tagged mutagenesis. *Infect. Immun.* **67**, 6385–6393 (1999).
- Miao, E. A. & Miller, S. I. A conserved amino acid sequence directing intracellular type III secretion by *Salmonella typhimurium*. *Proc. Natl Acad. Sci. USA* **97**, 7539–7544 (2000).
- Bakshi, C. S. *et al.* Identification of SopE2, a *Salmonella* secreted protein which is highly homologous to SopE and involved in bacterial invasion of epithelial cells. *J. Bacteriol.* **182**, 2341–2344 (2000).
- Kingsley, R. A. & Baumber, A. J. Host adaptation and the emergence of infectious disease: the *Salmonella* paradigm. *Mol. Microbiol.* **36**, 1006–1014 (2000).
- Sanderson, K. E. & Liu, S. L. Chromosomal rearrangements in enteric bacteria. *Electrophoresis* **19**, 569–572 (1998).
- Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474 (1997).
- Marcus, S. L., Brumell, J. H., Pfeifer, C. G. & Finlay, B. B. *Salmonella* pathogenicity islands: big virulence in small packages. *Microbes Infect.* **2**, 145–156 (2000).
- Perna, N. T. *et al.* Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* **409**, 529–533 (2001).
- Wong, K. K. *et al.* 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 (1998).
- Hashimoto, Y., Li, N., Yokoyama, H. & Ezaki, T. Complete nucleotide sequence and molecular characterization of ViB region encoding Vi antigen in *Salmonella typhi*. *J. Bacteriol.* **175**, 4456–4465 (1993).
- Mirol, S. *et al.* Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic *Salmonella typhimurium* strain. *Proc. Natl Acad. Sci. USA* **96**, 9845–9850 (1999).

- Zhang, X. L. *et al.* *Salmonella enterica* serovar typhi uses type IVB pili to enter human intestinal epithelial cells. *Infect. Immun.* **68**, 3067–3073 (2000).
- McClelland, M. *et al.* Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature*, **413**, 852–856 (2001).
- Sherburne, C. K. *et al.* The complete DNA sequence and analysis of R27, a large IncHI plasmid from *Salmonella typhi* that is temperature sensitive for transfer. *Nucleic Acids Res.* **28**, 2177–2186 (2000).
- Wain, J. *et al.* Quantitation of bacteria in blood of typhoid fever patients and relationship between counts and clinical features, transmissibility, and antibiotic resistance. *J. Clin. Microbiol.* **36**, 1683–1687 (1998).
- Hu, P. *et al.* Structural organization of virulence-associated plasmids of *Yersinia pestis*. *J. Bacteriol.* **180**, 5192–5202 (1998).
- Perry, R. D. & Fetherston, J. D. *Yersinia pestis*—etiologic agent of plague. *Clin. Microbiol. Rev.* **10**, 35–66 (1997).
- Prentice, M. B. *et al.* *Yersinia pestis* pFra shows biovar-specific differences and recent common ancestry with a *Salmonella enterica* serovar Typhi plasmid. *J. Bacteriol.* **183**, 2586–2594 (2001).
- Jing, D. H., Dong, F., Latham, G. J. & von Hippel, P. H. Interactions of bacteriophage T4-coded primase (gp61) with the T4 replication helicase (gp41) and DNA in primosome formation. *J. Biol. Chem.* **274**, 27287–27298 (1999).
- Doolittle, R. F., Feng, D. F., Tsang, S., Cho, G. & Little, E. Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science* **271**, 470–477 (1996).
- Parkhill, J. *et al.* Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* **404**, 502–506 (2000).
- Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944–945 (2000).

**Supplementary information** is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

## Acknowledgements

We would like to acknowledge the support of the Sanger Centre core sequencing and informatics groups. We are grateful to the *S. typhimurium* sequencing group at Washington University for sharing sequence and analysis data before publication. A.K. and T.S.L. were supported by a grant from the Danish National Research Foundation. This work was supported by the Wellcome Trust through its Beowulf Genomics initiative and by a Wellcome Trust Programme Grant to G.D.

Correspondence and requests for materials should be addressed to J.P. (e-mail: [parkhill@sanger.ac.uk](mailto:parkhill@sanger.ac.uk)) or G.D. (e-mail: [g.dougan@ic.ac.uk](mailto:g.dougan@ic.ac.uk)). The sequences have been submitted to EMBL with the accession numbers AL513382 (chromosome), AL513383 (pHCM1) and AL513384 (pHCM2).

# Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2

Michael McClelland\*, Kenneth E. Sanderson†, John Spieth‡, Sandra W. Clifton‡, Phil Latreille‡, Laura Courtney‡, Steffen Porwollik\*, Johar Ali‡, Mike Dante‡, Feiyu Du‡, Shunfang Hou‡, Dan Layman‡, Shawn Leonard‡, Christine Nguyen‡, Kelsi Scott‡, Andrea Holmes‡, Neenu Grewal‡, Elizabeth Mulvaney‡, Ellen Ryan‡, Hui Sun‡, Liliana Florea§, Webb Miller§, Tamberlyn Stoneking‡, Michael Nhan‡, Robert Waterston‡ & Richard K. Wilson‡

\* Sidney Kimmel Cancer Center, 10835 Altman Row, San Diego, California 92121, USA

† Department of Biological Sciences, University of Calgary, Calgary, Alberta T2N 1N4, Canada

‡ Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Boulevard, St Louis, Missouri 63108, USA

§ Department of Computer Science and Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

*Salmonella enterica* subspecies I, serovar Typhimurium (*S. typhimurium*), is a leading cause of human gastroenteritis, and is used as a mouse model of human typhoid fever<sup>1</sup>. The incidence of non-typhoid salmonellosis is increasing worldwide<sup>2–4</sup>, causing millions of infections and many deaths in the human population each year. Here we sequenced the 4,857-kilobase (kb) chromosome

Present address: Celera Genomics, Rockville, Maryland 20850, USA.