# Genomic Cleavage Map of *Salmonella typhi* Ty2

SHU-LIN LIU AND KENNETH E. SANDERSON\*

*Salmonella Genetic Stock Centre, Department of Biological Sciences, University of Calgary, Calgary, Alberta, Canada T2N 1N4*

Received 18 April 1995/Accepted 25 June 1995

**The genomic cleavage map of** *Salmonella typhi* **Ty2, 4,780 kb in size, was determined through digestion of the genomic DNA with endonucleases and separation of the fragments by pulsed-field gel electrophoresis. The chromosome has 33, 26, 7, and 35 sites for the enzymes** *Xba***I,** *Bln***I, I-***Ceu***I, and** *Spe***I, respectively. The fragments were arranged around the chromosome through excision of fragments from the gel, redigestion with a second enzyme, end labelling with 32P, and reelectrophoresis and named in alphabetical order. Tn***10* **transposons inserted in 82 different genes of** *Salmonella typhimurium* **were transduced by phage P22 into** *S. typhi***, and the location of Tn***10***, and thus of the gene, was mapped through the** *Xba***I and** *Bln***I sites of Tn***10***. All seven I-***Ceu***I sites (in** *rrl* **genes for 23S rRNA) were conserved, and the gene order within the I-***Ceu***I fragments resembles that of** *S. typhimurium* **LT2, but the order of I-***Ceu***I fragments is rearranged from ABCDEFG in** *S. typhimurium* **LT2 to AGCEFDB in** *S. typhi***. In addition, there is a 500-kb inversion which covers the terminus region. Comparisons of lengths of segments between genes showed that** *S. typhi* **has segments which differ in size from those in** *S. typhimurium***. The** *viaB* **locus, for synthesis of the Vi antigen of** *S. typhi***, was shown to be within a 118-kb loop (a segment of DNA with no homolog in most other** *Salmonella* **species) between** *mel* **and** *poxA* **on the chromosome.**

Of the more than 2,300 closely related *Salmonella* serovars recognized, *Salmonella typhi* is the only one that grows exclusively in humans and causes typhoid enteric fever  $(9, 20, 39)$ . Typhoid fever is a major public health problem in many parts of the world; 13 million cases per year are estimated, exclusive of China (9). Because of its importance as a human pathogen, *S. typhi* has been the target of extensive investigation (16, 35, 37, 48), including studies of its invasiveness of human epithelial cells (11, 13, 23).

The division of the genus *Salmonella* into many serovars is based on the somatic and flagellar antigens; these serovars have been called species (19), and *S. typhi* is one of the species. Molecular genetic evidence indicates close relationships of all of these serovars (8), and some researchers use a single species designation (21); in this report, we retain the use of different species names.

A few other *Salmonella* species, including *S. paratyphi* A, B, and C and *S. sendai*, are also human-adapted pathogens which cause similar but milder diseases, called paratyphoid fever. Reeves et al. (40), by using multilocus enzyme electrophoresis to measure the genetic distances within the seven *Salmonella* DNA hybridization subgroups, showed that *S. typhi* strains constitute a single clone which is widely separated from the other species in subgroup I. Selander et al. (47) also concluded that *S. typhi* is a relatively homogeneous group, with no close phylogenetic relationships to the paratyphoid agents.

Genetic changes which occur during evolution of bacteria can be classified into three types: (i) base pair changes, probably due primarily to replication errors, (ii) genomic rearrangements (inversions, translocations, and duplications) involving large blocks of DNA, and (iii) lateral transfer, which results in addition of new genetic material from an external source. Which of these is the most significant in the evolution of species of *Salmonella*? High levels of DNA reassociation and low thermal elution midpoint values in studies with differ-

ent species of the same genus indicate that base pair changes in homologous genes are infrequent (8). In addition, nucleotide sequencing of homologous genes of different species within subgenus I of *Salmonella* (which includes all of the human-pathogenic species usually studied) indicates low divergence (5), except in highly polymorphic genes, such as *rfb* (41), *fliC*, and *fljB* (22). Genomic rearrangements occur rarely in the genus *Salmonella*; *S. typhimurium* (27), *S. paratyphi* B (24), and *S. enteritidis* (26) have very similar genomic maps. Lateral transfer has resulted in species-specific loops of DNA: the *nanH* gene of *S. typhimurium* LT2 is absent from all other species of subgenus I (17); some *Escherichia coli* serovars have unique blocks of DNA called ''pathogenicity islands'' (4).

The objective of this study was to examine the ways *S. typhi* differs genetically from other *Salmonella* spp. We previously constructed genomic cleavage maps of *S. typhimurium* LT2 (27, 28), *S. enteritidis* (26), and *S. paratyphi* B (24); comparisons of these maps and that of *E. coli* (2, 38) show that these enteric bacteria share a common basic genomic structure, with some differences specific to individual bacteria. However, we recently reported that the *Xba*I, *Bln*I, and I-*Ceu*I genomic maps of strains of *S. typhi* show major rearrangements, including inversions and translocations due to homologous recombination in the *rrn* genes for rRNA (30). Here we present a detailed analysis of the genomic structure of *S. typhi*, including a *Spe*I map. Comparison with the genome of *S. typhimurium* shows several regions with excess DNA (loops of DNA) which are missing from most species of *Salmonella*. One of these loops is up to 118 kb long, contains the *viaB* locus for Vi antigen synthesis, and is not present in most other *Salmonella* species.

## **MATERIALS AND METHODS**

**Bacterial strains, phage, and cultivation conditions.** H251.1, a derivative of *S. typhi* Ty<sub>2</sub> which carries the mutation  $\Delta a \cdot \frac{1019}{9}$ , was obtained from D. M. Hone (15) and was the wild-type strain used to generate the endonuclease cleavage maps and to make Tn*10* insertion mutants. Independent *S. typhi* wild-type strains were obtained from several sources: 382-82, 9032-85, 1707-81, 1196-74, 3434-73, 3137-73, 9228-77, 3185-73, and Ty21a were from the Centers for Disease Control and Prevention, Atlanta, Ga. (40); SA4828 was from the Provincial Laboratory of

<sup>\*</sup> Corresponding author. Phone: (403) 220-6792. Fax: (403) 289- 9311. Electronic mail address: kesander@acs.ucalgary.ca.

Alberta, Calgary, Alberta, Canada; 25T-35, 25T-36, 25T-37, 25T-38, 25T-39, 25T-41, and a Ty2 strain were from the Laboratory Center for Disease Control, Ottawa, Ontario, Canada; IP E.88.374 and IP E.88.353 were from R. K. Selander (6); and ISP1820 and H238.2 were from D. M. Hone (15). Strains of other *Salmonella* species were from R. K. Selander. Cosmid pGBM124, containing a 14-kb DNA fragment of *S. typhi* including *viaB*, was obtained from T. Ezaki (14). The strains of *S. typhimurium* LT2 with Tn*10* insertions were part of a kit maintained at the *Salmonella* Genetic Stock Center (1, 45). Tn*10* was transduced from *S. typhimurium* LT2 to *S. typhi* H251.1 by P22 HT105/*int* followed by selection for tetracycline resistance. As observed earlier, the transduction frequency was low (10, 49). When the Tn*10* insertion in *S. typhimurium* produced a detectable phenotype, the transductant in *S. typhi* showed the same phenotype, indicating that insertion was due to homologous recombination.

Luria-Bertani medium (10 g of tryptone, 5 g of yeast extract, 10 g of NaCl, 3.5 ml of 1 M NaOH) was used for cultivation of all strains; solid medium also contained 1.5% agar. The minimal medium used is a modified Davis medium which has been described previously (46). For growth of strain H251.1, which carries the mutation  $\Delta a$ roC1019, 100  $\mu$ g of 2,3-dihydroxybenzoate per ml was added to Luria-Bertani medium; 100 mg each of 2,3-dihydroxybenzoate, *p*-aminobenzoate, *p*-hydroxybenzoate, phenylalanine, tyrosine, and tryptophan was added to minimal medium. Tetracycline was used at 20  $\mu$ g/ml. Strains were maintained in 15% glycerol at  $-70\degree C$ , and a single colony was isolated prior to use.

**Enzymes and chemicals.** Endonucleases were from New England Biolabs (*BlnI* [=AvrII], I-*CeuI*, and *SpeI*) and Boehringer Mannheim (*XbaI*). [<sup>32</sup>P]dCTP was from New England Nuclear. Most other chemicals, including agarose, were from Sigma Chemical Co.

**Pulsed-field gel electrophoresis (PFGE) methods.** Preparation of high-molecular-weight genomic DNA, endonuclease cleavage of DNA in agarose blocks, separation of DNA fragments by PFGE, and double-cleavage techniques were done as previously reported (27, 28).

**Southern hybridization.** Plasmid DNA was used as probes as previously described (28).

## **RESULTS**

*Xba***I, I-***Ceu***I,** *Bln***I, and** *Spe***I maps of** *S. typhi* **and their alignment with each other.** The genomic cleavage map of *S. typhi* Ty2, for the enzymes *Xba*I, *Bln*I, and I-*Ceu*I, contains 33, 26, and 7 fragments, respectively (Fig. 1) (30). The order of the fragments around the chromosome was determined for *Xba*I and *Bln*I, and the fragments were named in alphabetical order, starting with A (Fig. 1; Table 1). I-*Ceu*I, an intron-encoded endonuclease (34), digests a site in each of the seven *rrn* operons, producing seven fragments (25); each of the seven *S. typhi* fragments, A to G, carries genes homologous to the same fragment of *S. typhimurium* (Tables 1 and 2). In *S. typhimurium* LT2, these fragments are named in the order I-*Ceu*I-ABC DEFG (27), but in *S. typhi*, the order of these fragments around the chromosome is rearranged to I-*Ceu*I-AGCEFDB (30) (Fig. 1). Because of the rearrangements, the *rrn* operons are named as hybrids with the designations of *E. coli* and *S. typhimurium* in the I-*Ceu*I fragments postulated to be on their left and right; e.g., *rrnG/H* at position 300 kb is composed of *rrnG* on one end of I-*Ceu*I-B and *rrnH* on one end of I-*Ceu*I-A. Other *rrn* operons are named in the same way.

In *E. coli* K-12 and in the *Salmonella* species reported earlier (24, 26–28, 44), gene positions are designated in centisomes, or in kilobases, with *thr* genes at 0 U on the circular map. Arbitrarily, we show the large I-*Ceu*I A fragment in the same location and orientation as in *S. typhimurium*, with *proA* near position 500 kb and *nadB* near position 2600 kb; thus, I-*Ceu*I-G, which contains *thr* and is normally at 0 kb on the map, is now near position 3000 kb, and other genes are similarly displaced. The counterclockwise end of 2,400-kb I-*Ceu*I fragment A (at *rrnG/H*) is arbitrarily placed at position 300 kb (the distance of the *rrn* gene from *thr* in LT2, to align the main parts of the genomes for a more convenient comparison), and the clockwise (CW) end of the same fragment in *rrnH/G* is thus at position 2700 kb (Table 1). Other fragments are placed on the chromosome with reference to these points. The locations of the CW ends of *Xba*I, *Bln*I, and I-*Ceu*I fragments, in kilobases, are listed in Table 1.

The endonuclease *Spe*I, which recognizes the sequence AC TAGT, cleaved the genomic DNA of *S. typhi* Ty2 strain H251.1 to 35 fragments; 33 of these, from 770 to 12 kb, were visible on an ethidium bromide-stained gel (Fig. 2, lane 10), while two smaller fragments of 7.5 and 1 kb were visible only following end labelling of the fragments and autoradiography (data not shown). The 35 *Spe*I fragments were placed on the *Xba*I-*Bln*I-I-*Ceu*I map by two methods. First, since Tn*10* does not have a *Spe*I site, a *Spe*I fragment containing Tn*10* has an increased size (3.5 kb larger in the case of Tn*10* dTet and 9.3 kb larger in the case of wild-type Tn*10*), and this size increase was detectable on a PFGE gel (Fig. 2). (Digestion by *Xba*I or *Bln*I, which has recognition sites in Tn*10*, results in cleavage of an *Xba*I or *Bln*I fragment into two parts if that fragment contains a Tn*10* insertion; data on the sizes of these fragments were used to place genes on the *Xba*I and *Bln*I maps.) Lanes 1 to 9 of Fig. 2 show the fragments resulting from *Spe*I digestion and PFGE separation of DNAs from strains carrying wild-type Tn*10* inserted into known genes of *S. typhi*; in each case, a fragment can be identified by its increased size. For example, in lanes 4 and 5, containing DNAs from strains with Tn*10* insertions in genes *nadA* and *bio*, band I (195 kb) has increased in size to about 204 kb, indicating that Tn*10* is inserted in this fragment. *nadA* and *bio* are adjacent on the chromosomes of *S. typhimurium* (27) and *S. typhi* (30) at about 2 o'clock on I-*Ceu*I-A (Fig. 1); thus, *Spe*I-I is placed at this location. By using this system, we identified most of the *Spe*I fragments by analysis of 82 strains which carry different Tn*10* insertions (Table 2) and roughly located them on the *Xba*I, I-*Ceu*I, and *Bln*I maps of this strain (Fig. 1). Second, the exact locations of the cleavage sites of the 35 *Spe*I fragments were determined by double digestion performed as described earlier (27); briefly, agarose blocks containing individual fragments from a single digestion were excised from the gel, and the fragment was digested with the second enzyme, end labelled, and re-electrophoresed. For example, digestion of isolated *Xba*I-D with *Spe*I yielded three fragments, *Spe*I-9H (110 kb), *Spe*I-I (195 kb), and *Spe*I-J9 (35 kb); this shows that *Spe*I-I is contained within *Xba*I-D, with its counterclockwise end (at 707 kb on the map; Table 1) 110 kb inside the counterclockwise site in *Xba*I (at 597 kb) and its CW end (at 902 kb) 35 kb inside the CW *Xba*I site (at 937 kb). By the same methods, all *Xba*I, *Bln*I, I-*Ceu*I, and *Spe*I fragments larger than 100 kb were excised, digested separately with the three other enzymes, end labelled, and electrophoresed (data

FIG. 1. (A) Genomic cleavage map of *S. typhi* Ty2 for *Xba*I, I-*Ceu*I, *Bln*I, and *Spe*I. The sizes of the map in segments, in kilobases, are on the inside of the circle. For most enzymes, the fragments are named in alphabetical order, beginning with A at 0 kb; for I-*Ceu*I, the fragments are named to be homologous with those of other *Salmonella* spp. and *E. coli*, but the order of I-*Ceu*I fragments in *S. typhi* is rearranged to AGCEFDB (30). The enzyme I-*Ceu*I digests the seven *rrn* operons in the *rrl* gene for 23S rRNA (34). The structure of the *rrn* operons is shown in detail outside the circle; the positions of the cleavage sites in the *rrn* genes are from double-digestion data. The four *rrn* genes which have *Xba*I and *Bln*I sites are inferred to have the tRNA gene for glutamate tRNA, and the other three are inferred to have the gene for alanyl tRNA, by analogy with *E. coli* (36). The arrow indicates the direction of transcription originally determined for *rrnB* in *E. coli* (36). The relative locations of endonuclease cleavage sites (Table 1) were determined by double-digestion data (see text). The genes were located through mapping of the *Xba*I and *BlnI* sites in Tn*10* insertions in the genes (Table 2). The region which is inverted with respect to *S. typhimurium* is marked by bars at about 1,200 and 1,800 kb. (B) Structure of the *rrnB* operon of *E. coli* (36). Endonuclease cleavage sites were determined from nucleotide sequence, and the order of the genes is *rrs* (16S rRNA)-tRNA-*rrl* (23S rRNA)-*rrf* (5S rRNA). The nucleotide numbers of cleavage sites and of ends of genes are shown.







TABLE 1. *Xba*I, I-*Ceu*I, *Bln*I, and *Spe*I fragments from the genomic DNA of *S. typhi* Ty2

Frag- ment	XbaI		I-CeuI		Bln1		SpeI	
	Size $(kb)^a$		Site <sup>b</sup> Size $(kb)^a$	$\text{Site}^b$	Size $(kb)^{a}$	$\text{Site}^b$	Size $(kb)^a$	$\mathrm{Site}^b$
А	323	233	2,400	2,700	365	37	50	37
B	224	457	724	300	100	137	170	207
C	140	597	502	4,030	500	637	(31)	
D	340	937	136	4,356	32	669	(25)	
E	325	1,262	146	4,176	68	737	(14)	
F	140	1,402	44	4,220	830	1,567	56	333
G	275	1,677	828	3,528	265	1,832	104	437
H	34	1,711			32	1,864	270	707
I	306	2,017			23	1,887	195	902
$\bf J$	104	2,121			115	2,002	121	1,023
K	326	2,447			600	2,602	276	1,299
L	(34)				100	2,702	33	1,332
M	(5)	2,486			120	2,822	180	1,512
N	216	2,702			75	2,897	193	1,705
O	62	2,764			463	3,360	77	1,782
P	328	3,092			32	3,392	30	1,812
Q	(10)				138	3,530	80	1,892
R	(3)				280	3,810	(30)	
S	(1)				114	3,914	(1)	1,923
T	(0.5)	3,107			(100)		45	1,968
U	(3.2)				(3.5)	4,028	70	2,038
V	(2.5)	3,112			20	4,048	(52)	
W	160	3,272			78	4,126	(25)	
X	258	3,530			92	4,218	(12)	
Y	6.5	3,536			26	4,244	140	2,267
Ζ	20	3,556			208	4,452	380	2,647
AA	205	3,762					91	2,738
BB	82	3,844					7.3	2,745
CC	70	3,914					263	3,008
DD	114	4,028					99	3,107
EE	190	4,218					764	3,871
FF	420	4,638					392	4,263
GG	52	4,690					116	4,379
HH							95	4,474
$_{\rm II}$							293	4,767

*<sup>a</sup>* When the order of two or more fragments is not known, their sizes are in

 $\frac{b}{b}$  For *XbaI*, *BlnI*, and *SpeI*, location means the CW site, in kilobases, of each fragment on the genome from the arbitrary 0-kb point.

For more convenient comparison, the I-*CeuI* fragments are designated in accordance with those of *S. typhimurium* LT2 on the basis of homology.

not shown). These data revealed the locations of most cleavage sites, although the relative order of some smaller, adjacent fragments is not known (e.g., the sizes of *Spe*I-R and *Spe*I-S are in parentheses in Table 1 and are shown as a single space between *Spe*I-Q and *Spe*I-T in Fig. 1). The designations, sizes, and locations of the cleavage fragments resulting from digestion by the four endonucleases are in Table 1, and the maps are in Fig. 1. The fragments in which specific genes are located, on the basis of digestion of the DNAs of 82 strains containing Tn*10* insertions in specific genes, and their locations in kilobases around the chromosome of *S. typhi* Ty2 are in Table 2. Seventy-five genes were located in this way (seven insertions were not into known genes), and the seven *rrn* operons were located by determining the positions of I-*Ceu*I sites (Fig. 1).

**Comparison of the genomes of** *S. typhi* **and** *S. typhimurium.* To facilitate comparisons of the *S. typhi* and *S. typhimurium* genomes, we converted the circular maps into four vertical parts, aligned the homologous I-*Ceu*I fragments side by side in accordance with the *S. typhi* Ty2 order, which is AGCEFDB, and joined the positions of homologous genes by horizontal (or sloping) lines (Fig. 3). This analysis revealed three types of

genomic changes. Firstly, there are rearrangements of I-*Ceu*I fragments (discussed above and in reference 30). Secondly, there is an inversion of a segment of about 500 kb, including the *nar* gene (1,256 kb) at one end and the *pncX* gene (1,747 kb) at the other; this includes the terminus of replication (Ter). Thirdly, although gene order and interval length are very similar in the two species in most cases (e.g., *fliC-hisC-ack-purFptsI-purC*), there are differences in DNA segment length between some pairs of genes in the two species, suggesting the existence of insertions of unique DNA in one species or of deletions from the other; these excess segments are called loops.

The *viaB* locus, for synthesis of the Vi antigen of *S. typhi*, resides in one of these loops. Probing of *Xba*I and *Bln*I digests of *S. typhi* DNA with pGBM124 carrying a 14-kb fragment containing the *viaB* genes from *S. typhi* (14) showed that these genes are located on *Xba*I-X and *Bln*I-O (Fig. 4). This localizes the genes between 3,272 kb (the counterclockwise end of *Xba*I-X) and 3,370 kb (the CW end of *Bln*I-O), between the *mel* and *poxA* genes. This is consistent with genetic data which placed *viaB* adjacent to the *mel* and *inl* (18) genes. The 14-kb *viaB* gene probe detected homologous sequences in *S. typhi* and in *S. paratyphi* C, both of which express the Vi antigen, but not in *S. typhimurium* or in *S. paratyphi* A or B (data not shown). Thus, at least 14 kb, and possibly all, of the 118-kb loop of *S. typhi* DNA between *poxA* and *mel* was confirmed to be present in *S. typhi* but missing from *S. typhimurium* and most other *Salmonella* spp.

Analysis of Fig. 3 indicates the presence of several other loops. Interval length differences are seen between *pheA* and *rrnG/H* (at 270 kb on the *S. typhi* chromosome), *proBA* and *purE* (370 to 577 kb), *nadA* and *bio* (817 kb), *aspC* and *pncB* (1,007 kb), and *putA* and *flgL* (at 1,127 kb). The interval lengths between *pncB* and *pyrD* in *S. typhi* Ty2 (60 kb) and *S. typhimurium* LT2 (46 kb) are similar, but the same interval is only 15 kb in *S. enteritidis* and *S. paratyphi* B (24, 26). An insertion including the *nanH* gene (17) occurred between *pncB* and *pyrD* in *S. typhimurium* LT2 (27), but *S. typhi* lacks the *nanH* gene (17, 30a), so there must be a different (not *nanH*) and larger insertion between *pncB* and *pyrD* in *S. typhi* Ty2. I-*Ceu*I-D of *S. typhi* (at 4,300 kb) is 136 kb, while the homologous fragment of *S. typhimurium* (27) and other *Salmonella* species (26) is consistently only 92 kb, indicating a 44-kb loop in *S. typhi.*

*Xba*I digestion of the DNAs of 23 strains of *S. typhi*, most of which are independent wild-type strains, and ethidium bromide staining (Fig. 5A) or end labelling (Fig. 5B) to detect small fragments revealed that smaller *Xba*I fragments are strongly conserved in many strains while the larger *Xba*I fragments are quite variable. Of the six tiny *Xba*I fragments (Q to V) which constitute a cluster of unknown order at 3,100 kb on the map (Table 1), four of the fragments  $(Q, R, U,$  and V) are present in almost all of the 23 strains; the two smaller fragments, S and T, ran off this gel. *Xba*I-M (5 kb) and *Xba*I-Y (6.5 kb) are present in all of the strains. These small *Xba*I fragments were not seen in any of the many other *Salmonella* species tested (data not shown).

### **DISCUSSION**

The genome map of *S. typhi* Ty2 (Fig. 1) resembles those of *S. typhimurium*, *S. paratyphi* B, and *S. enteritidis*: genome sizes are all between 4,600 and 4,800 kb, the order of genes on chromosome segments is usually the same, and all have seven *rrn* operons, four of which have *Xba*I and *Bln*I sites, indicating the presence of the gene for glutamate tRNA.

TABLE 2. Locations of genes containing Tn*10* on the genome map of *S. typhi* H251.1

Gene with		Map position			
$_{\rm Tn10}$	XbaI	I-CeuI	BlnI	Spel	(kb)
relA21	А	B	Α	А	27
cysC1511	A	B	B	В	65
srl-202	А	B	C	в	151
proU1655	A	B	$\mathsf{C}$	B	172
tryrA555	B	в	C	$(56 \text{ kb})$	267
pheA534	B	B	C	(56 kb)	270
probA	B	A	C	G	369
zag-3030	B	A	$\overline{C}$	G	417
zag-208	в	А	С	G	422
purE884	C	A	C	Н	577
$ahp-11$	D	А	Е	Н	677
nadA213	D	A	F	I	817
bio-102	D	А	F	I	842
aspC409	E	A	F	J	1007
pncB150	Е	А	F	J	1017
pyrD2266	E	A	F	K	1077
putA810	E	А	F	K	1127
flgL	E	A	F	K	1132
pyrC	E	А	F	K	1172
ptsG	E	A	F	K	1212
chlC1130	E	А	F	K	1256
oppB255	F	A	F	K	1277
$trp-2451$	F	А	F	K	1292
$\cos B3305$	F	A	F	L	1303
pyrF696	F	А	F	L	1312
oxrA2	F	A	F	М	1370
ompD	G	Α	F	М	1455
tppB	G	А	G	N	1610
pncX	I	Α	G	О	1747
zea-1437	I	А	J	T	1937
cheY	I	Α	J	т	1952
cheB	I	А	J	т	1952
cheA	I	Α	J	T	1952
zea-4	I	А	J	т	1967
$\operatorname{fl}$ i $C$	I	Α	J	U	1977
$cob-2$	J	А	K	(52 kb)	2042
$\cosh-4$	J	Α	K	$(52 \text{ kb})$	2042
$\cosh 10$	J	A	K	$(52 \text{ kb})$	2042
hisG	J	Α	K		2088
hisD	J	А	K		2088
hisC8667	J	Α	K		2088
hisE	J	А	K		2088
ompC396	K	Α	K	Z	2312
ack-408	K	Α	K	Z	2391
<i>purF1714</i>	K	Α	K	Z	2421
ptsI421	N	Α	K	Z	2495
purC882	N	Α	K	Z	2546
purI175	N	Α	K	Z	2567
guaB	N	А	K	Z	2578
glyA540	N	А	L	AA	2651
purG1739	N	А	L	AA	2666
nadB214	N	А	L	AA	2685
pan-540	О	G	М	$_{\rm CC}$	2753
leu-1151	P	G	N	<b>CC</b>	2853
pyrA234	P	G	О	CC	2908
thr-577	$\mathbf P$	G	О	CC	2983
serB965	P	G	О	CC	2998
pyrB692	W	G	О	EΕ	3140
purA874	W	G	О	EE	3216
mutL111	W	G	О	ЕE	3221
poxA401	W	G	О	EΕ	3234
mel-351	X	G	O	ЕE	3405
met-900	X	G	О	EΕ	3525
crp-773	AA	C	R	EE	3564
cysG1510	AA	$\mathcal{C}$	R	EΕ	3579

TABLE 2—*Continued*

Gene with		Map position				
Tn10	$I-CeuI$ XbaI		Bln <sub>I</sub>	SpeI	(kb)	
xyl-183	ΒB	C	R	EΕ	3785	
uncA	DD	C	$(100 \text{ kb})$	FF	4003	
metB879	EE	Е	X	FF	4139	
metF	EЕ	E	X	FF	4148	
ppc-2	EE	Е	X	FF	4158	
argH1823	EЕ	Е	X	FF	4165	
ilvA595	FF	D	Ζ	GG	4273	
ilvE	FF	D	Z	GG	4274	
cya-1091	FF	D	Z	GG	4300	
$m$ et $\emph{E}$	FF	D	Z	GG	4320	
ntrA	FF	в	Ζ	HН	4419	
argG	FF	в	А	HН	4455	
metC975	FF	в	A	П	4598	
serA977	GG	B	A	Н	4684	
pepP	GG	в	А	П	4688	
lysA565	А	в	А	П	4735	
argA1832	A	в	А	П	4765	

However, there are also important differences between the *S. typhi* genome and those of other *Salmonella* species. The chromosome of *S. typhi* shows two categories of genetic rearrangements. Inversions and translocations postulated to be due



FIG. 2. PFGE of SpeI-cleaved genomic DNA of S. typhi Ty2 strain H251.1 with a Tn10 insertion transduced from S. typhimurium LT2. Lanes: 1, proBA:: Tn10; 3, purE::Tn10; 3, ahp::Tn10; 3, thadA::Tn10; 5, bio::Tn10; 6, aspC:: are shown on the left, and the lanes are indicated; the designations and sizes of the *Spe*I fragments of the wild-type Ty2 strain and the sizes of the lambda molecular size markers, in kilobases, are on the right.

*Continued*



FIG. 3. Comparison of genome maps of 8. *ophi 1*Ty2 and 8. *ophimurium* LT2. The map of Ty2, redrawn from Fig. 1, is divided into four vertical segments starting at 0 kb (from Fig. 1) and shows the positions of Xbal (X), side. Homologous genes of S. pphi and S. pphimurium are joined by a horizontal (or sloping) line. Similar lines join XbaI sites if they are inferred from location to be in homologous sites.



FIG. 4. Localization of the *viaB* locus of *S. typhi* on the genomic cleavage map. (A) PFGE gel of genomic DNA of *S. typhi* Ty2 cleaved with *Bln*I (lane 1) and *Xba*I (lane 2). (B) Autoradiograph of panel A probed with pGBM124, which has a 14-kb insert of *S. typhi* DNA containing the *viaB* locus (14).

to homologous recombination between the *rrn* genes result in rearrangement of the genome of *S. typhi* Ty2 (Fig. 1); these are discussed in reference 30.

The second category of genetic rearrangement is an inversion, in *S. typhi* compared with *S. typhimurium*, of a segment of about 500 kb which includes the terminus region (Ter) (Fig. 1 and 6). This is the only region in which inversion in wild-type *Salmonella* strains is common. Inversions covering the Ter region, compared with *S. typhimurium*, are known in the following species: *E. coli*, 480 kb (7, 43); *S. enteritidis*, 750 kb (26); and *S. typhi*, 500 kb (this report) (Fig. 6). No inversion was

detected in *S. paratyphi* B (24). All of these inversions differ by at least one end point. The gene order in *S. typhimurium* may be considered the ancestral order, because all inverted orders could be obtained in single events from the *S. typhimurium* order, whereas more complex events would be required from other orders, but this may simply reflect a high frequency of inversions in this region. Inversions have been detected in experimental studies in *E. coli* K-12 and *S. typhimurium* LT2, although some end points are much less common than others (33). Two types of recombination occur at high frequency in *E. coli* K-12 in the Ter region: *recA*-dependent homologous recombination, resulting in deletions, and site-specific recombination at the *dif* site in the Ter region (31, 32). The role of these recombination events in inversion in wild-type strains is unknown. Surprisingly, in experimental studies with *E. coli* K-12, inversions with end points in the regions just outside TerA and TerC, called nondivisible zones (NDZ in Fig. 6), are seldom found, but in natural *Salmonella* and *E. coli* populations, this is the only region (except for *rrn* genes in *S. typhi*) in which inversions are common (Fig. 6); the rules governing survival of such inversions in natural populations must differ from laboratory studies in some undefined way.

Even though the gene orders in their chromosomes are very similar, the numbers and locations of *Xba*I and *Bln*I sites in the chromosomes are very different in *Salmonella* species, for *S. typhi* (this report), *S. typhimurium* (27), and *S. enteritidis* (26) have 33, 24, and 16 *Xba*I sites and 26, 16, and 12 *Bln*I sites, respectively. Thus, fingerprints of these species with these enzymes suggest major differences, yet the gene orders are very similar. As expected, I-*Ceu*I sites in the *rrn* operons and *Xba*I and *Bln*I sites in the glutamate tRNA genes in the *rrn* operons are fully conserved in all of these species, but the *Bln*I site in *rrs* (16S rRNA) genes of some *Salmonella* species (as in *E. coli rrnB* [Fig. 1B]) are missing from all seven *rrn* operons of *S. typhi*.

Rearrangements (inversions and translocations) may be a compensation mechanism to maintain genomic balance, keep-



FIG. 5. PFGE gel of XbaI-cleaved genomic DNAs of wild-type S. typhi strains. The cleavage products were end labelled with [<sup>32</sup>P]CTP and the Klenow fragment of DNA polymerase I before the run. Lanes: 1, 382-82; 2, 9032-85; for Disease Control and Prevention); 11, Ty2 (D. M. Hone); 12, H251.1; 13, lambda *Hin*dIII size markers; 14, Ty2 (Laboratory Center for Disease Control); 15, 25T-35; 16, 25T-36; 17, 25T-37; 18, 25T-38; 19, 25T-39; 20, IP E.88.374; 21, 25T-41; 22, IP E.88.353; 23, ISP1820; 24, H238.2. (A) Ethidium bromide staining of the PFGE gel. (B) Autoradiograph of panel A to show the bands smaller than 10 kb. The *Xha*I fragments and their sizes in kilobases are on the left in panel B. The positions of fragments from  $\hat{H}$ indIII digestion of  $\lambda$  DNA are on the right.



FIG. 6. Inversions covering the terminus region in *Salmonella* spp. and *E. coli*. The vertical bar and genes on the left (S. tm) is the order of genes in *S. typhimurium* LT2, with the positions of these genes shown in kilobases (44). The closed bars to the right indicate segments of the chromosomes of *S. enteritidis* (S. en) (26), *S. typhi* (S. ty) (this report), and *E. coli* K-12 (E. co) (2, 38) which are inverted. TerA (*=pslA*) and TerC (*=plrA*) indicate the locations of replication termination in *E. coli* K-12 (12, 31, 32, 42). NDZ (nondivisible zones) indicates regions in which endpoints of inversions did not occur in experimental studies with *E. coli* K-12 (12, 42).

ing the origin (*oriC*) and termination (Ter) of replication separated by 180°, after an insertion of DNA has produced an imbalance. Thus, for example, the insertion of 118 kb containing the *viaB* locus between the *poxA* and *mel* genes (Fig. 1, 3, and 4), which would disrupt the genomic balance, might have triggered inversions or other rearrangements which would restore the balance.

How does *S. typhi* differ genetically from other *Salmonella* spp., and which of these changes may be the basis for the unique phenotypes, such as adaptation to humans and the capacity to cause typhoid enteric fever? Base pair changes in homologous genes are infrequent, for reassociation of DNA from *S. typhimurium* and *S. typhi* at 60°C is 88% and the thermal elution midpoint is only  $1.1^{\circ}C$  (8). Genomic rearrangements (due to homologous recombination in *rrn* genes and inversion in the Ter region) are surprisingly common, but it is not clear if these have a role in modifying the phenotype. Loops of *S. typhi*-unique DNA, probably due to lateral transfer, are common; several are detected as longer gene segments in *S. typhi* than in *S. typhimurium* (Fig. 3), and one between *mel* and *poxA* was proven to carry the *viaB* locus (Fig. 4). In addition, I-*Ceu*I-D, which is about 92 kb in all of the other serovars examined, is 136 kb in *S. typhi* (at position 4300 kb in Fig. 3), indicating an insertion of about 44 kb in this region. *S. typhimurium* has recently been shown to have an insertion, in the interval near *glyS* at position 3800 kb (Fig. 3), of about 8 kb of DNA containing the *lpfABCDE* genes for fimbriae; this DNA is missing from *S. typhi* and many other species (3). We did not detect this insertion by our methods, possibly because of low resolution in this area because there are no Tn*10* insertions in the 200-kb interval between *cysG* and *xyl* (Fig. 3). Lateral transfer resulting in loops of *S. typhi*-specific DNA is probably the most important genetic mechanism resulting in adaptation to humans and causation of enteric fever.

We reported previously that not only is the order of DNA segments on the chromosome of *S. typhi* Ty2 different from that of *S. typhimurium* LT2, but the orders of DNA segments on the chromosomes of different wild-type strains of *S. typhi* frequently differ from each other, apparently because of homologous recombination between *rrn* genes (30). Digestion by endonucleases such as *Xba*I do not efficiently reveal rearrangements, although they show considerable variability in the fingerprint patterns of different *S. typhi* strains (Fig. 5); however, rearrangements between the *rrn* genes are efficiently revealed by digestion with I-*Ceu*I, an endonuclease which digests the DNA only in the *rm* genes (25). We have evidence that such rearrangements are also common in *S. paratyphi* A (30a) and C (14a) but are not common in *S. typhimurium* strains (29).

## **ACKNOWLEDGMENTS**

We greatly appreciate the technical assistance of Lili Lei and Ning Qi. Andrew Hessel made helpful comments on the manuscript. We thank David Hone for providing *S. typhi* H251.1.

This work was supported by an operating grant from the Natural Sciences and Engineering Research Council of Canada and by grant RO1AI34829 from the National Institute of Allergy and Infectious Diseases.

#### **REFERENCES**

- 1. **Altman, E., J. R. Roth, A. Hessel, and K. E. Sanderson.** 1995. Transposons in current use in genetic analysis in *Salmonella. In* F. C. Neidhardt, C. A. Gross, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. Reznikoff, M. Riley, and M. Schaechter (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, 2nd ed., in press. American Society for Microbiology, Washington, D.C.
- 2. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54:**130–197.
- 3. **Baumler, A. J., and F. Heffron.** 1995. Identification and sequence analysis of *lpfABCDE*, a putative fimbrial operon of *Salmonella typhimurium*. J. Bacteriol. **177:**2087–2097.
- 4. **Blum, G., M. Ott, A. Lischewski, A. Ritter, H. Imrich, H. Tschape, and J. Hacker.** 1994. Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. Infect. Immun. **62:**606–614.
- 5. **Boyd, E. F., K. Nelson, F. S. Wang, T. S. Whittam, and R. K. Selander.** 1994. Molecular genetic basis of allelic polymorphism in malate dehydrogenase (*mdh*) in natural populations of *Escherichia coli* and *Salmonella enterica*. Proc. Natl. Acad. Sci. USA **91:**1280–1284.
- 6. **Boyd, E. F., F. S. Wand, P. Beltran, S. A. Plock, K. Nelson, and R. K. Selander.** 1993. *Salmonella* reference collection B (SARB): strains of 37 serovars of subspecies I. J. Gen. Microbiol. **139:**1125–1132.
- 7. **Casse, F., M.-C. Pascal, and M. Chippaux.** 1972. Comparison between the chromosome maps of *Escherichia coli* and *Salmonella typhimurium*: length of the inverted segment in the *trp* region. Mol. Gen. Genet. **51:**253–275.
- 8. **Crosa, J. H., D. J. Brenner, W. H. Ewing, and S. Falkow.** 1973. Molecular

relationships among the salmonelleae. J. Bacteriol. **115:**307–315.

- 9. **Edelman, R., and M. M. Levine.** 1986. Summary of an international workshop on typhoid fever. Rev. Infect. Dis. **8:**329–349.
- 10. **Edwards, M. F., and B. A. D. Stocker.** 1988. Construction of ΔaroA his Δpur
- strains of *Salmonella typhi*. J. Bacteriol. **170:**3991–3995. 11. **Elsinghorst, E. A., L. S. Baron, and D. J. Kopecko.** 1989. Penetration of human intestinal epithelial cells by *Salmonella*: molecular cloning and expression of *Salmonella typhi* invasion determinants in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **86:**5173–5177.
- 12. **Francois, V., J. Louarn, J. Patte, J.-E. Rebollo, and J.-M. Louarn.** 1990. Constraints in chromosomal inversions in *Escherichia coli* are not explained by replication pausing at inverted terminator-like sequences. Mol. Microbiol. **4:**537–542.
- 13. **Galan, J. E., and R. Curtiss.** 1991. Distribution of the *invA*, *-B*, *-C*, and *-D* genes of *Salmonella typhimurium* among other *Salmonella* serovars: *invA* mutants of *Salmonella typhi* are deficient for entry into mammalian cells. Infect. Immun. **59:**2901–2908.
- 14. **Hashimoto, Y., T. Ezaki, N. Li, and H. Yamamoto.** 1991. Molecular cloning of the *viaB* region of *Salmonella typhi*. FEMS Microbiol. Lett. **90:**53–56. 14a.**Hessel, A., and K. E. Sanderson.** Unpublished data.
- 15. **Hone, D. M., A. M. Harris, S. Chatfield, G. Dougan, and M. M. Levine.** 1991. Construction of genetically defined double *aro* mutants of *Salmonella typhi*. Vaccine **9:**810–815.
- 16. **Hook, E. W.** 1979. *Salmonella* species (including typhoid fever), p. 1256– 1269. *In* G. L. Mandell, R. G. Douglas, Jr., and J. E. Bennett (ed.), Principles and practices of infectious diseases. John Wiley & Sons, Inc., New York.
- 17. **Hoyer, L. L., A. C. Hamilton, S. M. Steenbergen, and E. R. Vimr.** 1992. Cloning, sequencing, and distribution of the *Salmonella typhimurium* LT2 sialidase gene, nanH, provides evidence for interspecies gene transfer. Mol. Microbiol. **6:**873–884.
- 18. **Johnson, E. M., B. Krauskopf, and L. S. Baron.** 1965. Genetic mapping of Vi and somatic antigen determinants in *Salmonella*. J. Bacteriol. **90:**302–308.
- 19. **Kauffman, F.** 1966. The bacteriology of Enterobacteriaceae. The Williams & Wilkins Co., Baltimore.
- 20. **Le Minor, L.** 1988. Typing of *Salmonella* species. Eur. J. Clin. Microbiol. Infect. Dis. **7:**214–218.
- 21. **Le Minor, L., and M. Y. Popoff.** 1987. Designation of *Salmonella enterica* sp. nov. as the type and only species of the genus *Salmonella*. Int. J. Syst. Bacteriol. **37:**465–468.
- 22. **Li, J., K. Nelson, A. C. McWhorter, T. S. Whittam, and R. K. Selander.** 1994. Recombinational basis of serovar diversity in Salmonella enterica. Proc. Natl. Acad. Sci. USA **91:**2552–2556.
- 23. **Liu, S.-L., T. Ezaki, H. Miura, K. Matsui, and E. Yabuuchi.** 1988. Intact motility as a *Salmonella* invasion-related factor. Infect. Immun. **56:**1967– 1973.
- 24. **Liu, S.-L., A. Hessel, H.-Y. M. Cheng, and K. E. Sanderson.** 1994. The *Xba*I-*Bln*I-*Ceu*I genomic cleavage map of *Salmonella paratyphi* B. J. Bacteriol. **176:**1014–1024.
- 25. **Liu, S.-L., A. Hessel, and K. E. Sanderson.** 1993. Genomic mapping with I-*Ceu*I, an intron-encoded endonuclease, specific for genes for ribosomal RNA, in *Salmonella* spp., *Escherichia coli*, and other bacteria. Proc. Natl. Acad. Sci. USA **90:**6874–6878.
- 26. **Liu, S.-L., A. Hessel, and K. E. Sanderson.** 1993. The *Xba*I-*Bln*I-*Ceu*I genomic cleavage map of *Salmonella enteritidis* shows an inversion relative to *Salmonella typhimurium* LT2. Mol. Microbiol. **10:**655–664.
- 27. **Liu, S.-L., A. Hessel, and K. E. Sanderson.** 1993. The *Xba*I-*Bln*I-*Ceu*I genomic cleavage map of *Salmonella typhimurium* LT2 determined by double digestion, end labelling, and pulsed-field gel electrophoresis. J. Bacteriol. **175:**4104–4120.
- 28. **Liu, S.-L., and K. E. Sanderson.** 1992. A physical map of the *Salmonella typhimurium* LT2 genome made by using *Xba*I analysis. J. Bacteriol. **174:** 1662–1672.
- 29. **Liu, S.-L., and K. E. Sanderson.** 1995. I-*Ceu*I reveals conservation of the genome of independent strains of *Salmonella typhimurium*. J. Bacteriol. **177:**3355–3357.
- 30. **Liu, S.-L., and K. E. Sanderson.** 1995. Rearrangements in the genome of the bacterium *Salmonella typhi*. Proc. Natl. Acad. Sci. USA **92:**1018–1022.
- 30a.**Liu, S.-L., and K. E. Sanderson.** Unpublished data.
- 31. Louarn, J., F. Cornet, V. François, J. Patte, and J.-M. Louarn. 1994. Hyperrecombination in the terminus region of the *Escherichia coli* chromosome: possible relation to nucleoid organization. J. Bacteriol. **176:**7524– 7531.
- 32. **Louarn, J.-M., J. Louarn, V. Francois, J. Patte.** 1991. Analysis and possible role of hyperrecombination in the termination region of the *Escherichia coli* chromosome. J. Bacteriol. **173:**5097–5104.
- 33. **Mahan, M. J., A. M. Segall, and J. R. Roth.** 1990. Recombination events that rearrange the chromosome: barriers to inversion, p. 341–349. *In* K. Drlica and M. Riley (ed.), The bacterial chromosome. American Society for Microbiology, Washington, D.C.
- 34. **Marshall, P., and C. Lemieux.** 1991. Cleavage pattern of the homing endonuclease encoded by the fifth intron in the chloroplast subunit rRNA-encoding gene of *Chlamydomonas eugamatos*. Gene **104:**1241–1245.
- 35. **Moshitch, S., L. Doll, B. Z. Rubinfeld, B. A. D. Stocker, G. K. Schoolnik, Y. Gafni, and G. Frankel.** 1992. Mono- and bi-phasic *Salmonella typhi*: genetic homogeneity and distinguishing characteristics. Mol. Microbiol. **6:**2589– 2597.
- 36. **Noller, H. F., and M. Nomura.** 1987. Ribosomes, p. 104–125. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington,  $D C$
- 37. **Pang, T., M. Altwegg, G. Martinetti, C.-L. Koh, and S. Puthucheary.** 1992. Genetic variation among Malaysian isolates of *Salmonella typhi* as detected by ribosomal RNA gene restriction patterns. Microbiol. Immunol. **36:**539– 543.
- 38. **Perkins, J. P., J. D. Heath, B. R. Sharma, and G. M. Weinstock.** 1994. *Xba*I and *Bln*I genomic cleavage maps of *Escherichia coli* K-12 strain MG1655 and comparative analysis of other strains. J. Mol. Biol. **232:**419–445.
- 39. **Popoff, M. Y., J. Bockemuel, and A. McWhorter-Murlin.** 1993. Supplement 1992 (no. 36) to the Kauffmann-White scheme. Res. Microbiol. **144:**495–498.
- 40. **Reeves, M. W., G. M. Evins, A. A. Heiba, B. D. Plikaytis, and J. J. Farmer III.** 1989. 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:**313–320.
- 41. **Reeves, P. R.** 1993. Evolution of *Salmonella* O antigen variation by interspecific gene transfer on a large scale. Trends Genet. **9:**17–22.
- 42. **Regello, J. E., V. Francois, and J.-M. Louarn.** 1988. Detection and possible role of two large nondivisible zones on the *Escherichia coli* chromosome. Proc. Natl. Acad. Sci. USA **85:**9391–9395.
- 43. **Sanderson, K. E., and C. A. Hall.** 1970. F-prime factors of *Salmonella* and an inversion between *S. typhimurium* and *Escherichia coli*. Genetics **64:**214–228.
- 44. **Sanderson, K. E., A. Hessel, and K. E. Rudd.** 1995. Genetic map of *Salmonella typhimurium*, edition VIII. Microbiol. Rev. **59:**241–303.
- 45. **Sanderson, K. E., and J. R. Roth.** 1988. Linkage map of *Salmonella typhimurium*, edition VII. Microbiol. Rev. **52:**485–532.
- 46. **Sanderson, K. E., and B. A. D. Stocker.** 1981. Gene *rfaH*, which affects lipopolysaccharide core structure in *Salmonella typhimurium*, is required also for expression of F factor functions. J. Bacteriol. **146:**535–541.
- 47. **Selander, R. K., P. Beltran, N. H. Smith, R. Helmuth, F. A. Rubin, D. J. Kopecko, K. Ferris, B. D. Tall, A. Cravioto, and J. M. Musser.** 1990. Evolutionary genetic relationships of clones of *Salmonella* serovars that cause human typhoid and other enteric fevers. Infect. Immun. **58:**2262–2275.
- 48. **Thong, K.-L., Y.-M. Cheong, S. Puthucheary, C.-L. Koh, and T. Pang.** 1994. Epidemiologic analysis of sporadic *Salmonella typhi* isolates and those from outbreaks by pulsed-field gel electrophoresis. J. Clin. Microbiol. **32:**1135– 1141.
- 49. **Zahrt, T. C., G. C. Mora, and S. Maloy.** 1994. Inactivation of mismatch repair overcomes the barrier to transduction between *Salmonella typhimurium* and *Salmonella typhi*. J. Bacteriol. **176:**1527–1529.