

Adaptation of Multilocus Sequencing for Studying Variation Within a Major Clone: Evolutionary Relationships of *Salmonella enterica* Serovar Typhimurium

Honghua Hu,* Ruiting Lan[†] and Peter R. Reeves*¹

*School of Molecular and Microbial Biosciences, The University of Sydney, Sydney, New South Wales 2006, Australia and [†]School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, New South Wales 2052, Australia

Manuscript received June 3, 2005

Accepted for publication September 21, 2005

ABSTRACT

Serovar Typhimurium of *Salmonella enterica* is a model organism for studies of pathogenesis that exhibits phage-type variation and variation in host range and virulence, but in a recent study showed no sequence variation in four genes, indicating the clonal nature of this serovar. We determined the relationships of 46 Typhimurium isolates of nine phage types using mutational changes detected either by matching AFLP (amplified fragment length polymorphism) fragments to computer-modeled LT2 AFLP fragments or by sequencing intergenic regions. Fifty-one polymorphic sites were detected, which gave a single phylogenetic tree. Comparison with genome sequences of five other serovars, Typhi, Paratyphi A, Gallinarum, Enteritidis, and Pullorum, enabled determination of the root of the tree. Only two parallel events were observed, giving high confidence in the tree branching order. The mutation-based tree provided a high level of consistency and a clear lineage for the Typhimurium isolates studied. This enabled us to show that for seven of the nine phage types used, the isolates studied have a single origin, but that two phage types clearly have more than one independent origin. We found that sequencing intergenic regions provides a good strategy for detection of mutational polymorphisms and study of phylogenetic relationships of closely related isolates and would be applicable to many other species.

SALMONELLA *enterica* serovar Typhimurium (denoted Typhimurium in this article) is a common cause of salmonellosis among humans and domestic animals worldwide. Most Typhimurium strains have a broad host range, causing diseases in multiple species, including humans, farm livestock, domestic fowl, rodents, and birds, whereas some Typhimurium variants have a very narrow host range, such as pigeon-adapted variants (BÄUMLER *et al.* 1998; RABSCH *et al.* 2002).

The Anderson phage-typing scheme (ANDERSON *et al.* 1977) is commonly used in epidemiological surveillance of Typhimurium infections and provides a means of following the rise and fall of different forms and geographical distribution of Typhimurium strains. A few phage types tend to dominate within a geographical region for a significant period of time. For example, multidrug-resistant DT104 has become a widespread pathogen in humans and animals in Europe and the United States since 1994 (LOW *et al.* 1997; GLYNN *et al.* 1998), while remaining rare in Australia. DT9 and DT135 are dominant in Typhimurium infections in Australia (POWLING 1997a,b), but DT9 is rare in Europe and the United States. The availability of a phage-typing

scheme for Typhimurium, and the wealth of data generated using it, make Typhimurium an excellent model for studying variation within a major bacterial clone.

We previously reported (HU *et al.* 2002) the use of amplified fragment length polymorphism (AFLP; VOS *et al.* 1995) to explore the relationships of nine phage types of Typhimurium. The dendrogram based on presence or absence of AFLP fragments showed good correlation with phage type, grouping most isolates of the same phage type together, demonstrating potential for development of a PCR- or microarray-based typing scheme.

However, the variation was mostly due to gain or loss of mobile DNA elements, which overwhelmed mutational variation in the AFLP tree (HU *et al.* 2002). Mutational changes, which accumulate over time, would better reflect phylogenetic relationships of Typhimurium isolates than polymorphisms relating to mobile elements, such as phages and plasmids. In this article, we focus on mutational differences to reveal the phylogenetic relationships of 46 isolates of nine phage types used in our previous study (HU *et al.* 2002).

Because the isolates in this study are within a major clone and closely related, mutational changes are expected to be rare, and a recent study (FAKHR *et al.* 2005) confirmed this with no variation among 85 Typhimurium isolates in a four-gene multilocus sequence typing (MLST) study, although MLST gave good discriminatory power for *S. enterica* as a whole (KOTETISHVILI *et al.* 2002).

Sequences data from this article have been deposited with the GenBank Data Libraries under accession nos. DQ077505–DQ077552.

¹Corresponding author: School of Molecular and Microbial Biosciences, The University of Sydney, Sydney, New South Wales 2006, Australia. E-mail: reeves@angis.usyd.edu.au

Genome sequence comparisons can be used to identify mutational changes in closely related isolates where more than one sequence is available. This approach was used to explore the evolutionary relationships of closely related isolates in *Bacillus anthracis* (READ *et al.* 2002) and *Mycobacterium tuberculosis* (GUTACKER *et al.* 2002; ALLAND *et al.* 2003). However, a comparative genome-based analysis could not be applied to Typhimurium, as the sequence of only one Typhimurium genome (LT2) was available.

In this study, we used two approaches to identify mutational polymorphisms among closely related isolates of a clone with only one genome sequence available. First, we identified mutation-related AFLP fragments by matching observed AFLP fragments with computer-modeled AFLP fragments to identify variation due to mutational changes. Second, we sequenced intergenic regions (IGs), which were relatively rich in mutational polymorphisms, and used PCR–restriction fragment length polymorphism (RFLP) and SNaPshot to investigate allele distribution in all isolates studied. We were able to determine the phylogenetic relationships for 46 Typhimurium isolates essentially on the basis of mutational changes.

MATERIALS AND METHODS

Bacterial isolates: Forty-six Typhimurium isolates from nine phage types, DT9, DT135, DT64, DT44, DT126, DT12a, DT1, DT141, and DT108, were used in this study. The isolates were from different sources, including human, animal, and environment, and from different regions of Australia. Details were given in Hu *et al.* (2002).

Primers: Primers for PCR and sequencing of AFLP fragments and IGs are in supplemental Tables S1 and S2, respectively (at <http://www.genetics.org/supplemental/>), and were based on the published Typhimurium LT2 genome sequence (McCLELLAND *et al.* 2001) (GenBank accession no. AE006468). Extension primers used for SNaPshot reaction (supplemental Table S3 at <http://www.genetics.org/supplemental/>) were reverse-phase cartridge purified. All primers were synthesized by SIGMA.

Computer modeling of AFLP based on LT2 genome sequence: The complete genome sequence of Typhimurium LT2 was computer modeled to emulate *MseI* and *EcoRI* digestion. Each of the 16 subsets for *MseI* +1/*EcoRI* +1 primer pair combinations was selected from the fragments generated. The sizes of fragments for each subset and their LT2 genome locations were imported into spreadsheets and the fragment sizes adjusted for addition of AFLP primers.

PCR and sequencing: PCR was normally carried out in 100- μ l volumes, which contained 1 μ l of DNA template, 1 \times PCR buffer, 0.2 mM of each dNTP, 250 nM of each primer, 2.5 units of Taq polymerase, and 20 μ g of BSA. PCR fragments were amplified in a PC-960G thermal cycler (Corbett Research, Australia) under the following conditions: initial denaturation at 94 $^{\circ}$ for 2 min, followed by 30 cycles of 94 $^{\circ}$ for 15 sec, 48–60 $^{\circ}$ for 30 sec, and 72 $^{\circ}$ for 1–2 min, final extension at 72 $^{\circ}$ for 15 min. PCR of IGs for PCR–RFLP and SNaPshot templates was performed under the same thermocycling conditions as above except that the reaction volume was 20 μ l.

PCR products were purified using UltraClean PCR clean-up kit (Geneworks). The procedures were carried out following

the manufacturer's instructions. Sequencing was carried out using an ABI 377 automated DNA sequencing system and ABI Dye Terminator Cycle sequencing kit (Perkin-Elmer).

Sequence analysis: DNA sequences were assembled and edited using Phred, Phrap, and Consed (GORDON *et al.* 1998). Multiple sequence alignments and comparisons were made using ClustalW (THOMPSON *et al.* 1994) and MULTICOMP (REEVES *et al.* 1994). MULTICOMP gives pairwise percentage differences of DNA and derived amino acid sequences to check for synonymous and nonsynonymous substitutions in coding regions. Nip (nucleotide interpretation program) (STADEN 1994) was used to search for recognition sites of restriction enzymes.

PCR–RFLP: IGs were amplified by PCR. A total of 15 μ l of PCR product was then digested with 10 units of one of the following restriction enzymes at 37 $^{\circ}$ for 4 hr: *HaeIII*, *HinfI* (Roche); *AccI*, *BspHI*, *Sau96I* (New England Biolabs); *Alu26I*, *Cfr10I*, *MvaI*, *TaaI*, *TaqI*, *VspI* (MBI Fermentas); *CfoI*, *HaeII* (Promega). These restriction enzymes were specifically chosen for the effect of mutational changes, resulting in one allele being digested, but not the other. The digested DNA fragments were separated electrophoretically through 2% agarose gel and the restriction digestion patterns were visually scored. The restriction enzymes for each single nucleotide polymorphism (SNP) and the resultant RFLP patterns are given in supplemental Table S4 (<http://www.genetics.org/supplemental/>).

SNaPshot: SNaPshot is based on the dideoxy single-base extension of oligonucleotide primers to interrogate and differentiate known SNPs. In this study, multiplex primer extension was performed to detect three or four SNPs in a single SNaPshot reaction. The extension primers in any reaction group differed by at least 4 bp, ranging from 19 to 34 bp. Supplemental Table S3 (<http://www.genetics.org/supplemental/>) shows the list of SNaPshot extension primers in groups and expected extension base for each polymorphism. The concentration of PCR products for each SNP was determined, and Milli-Q water was added to give a concentration of 2 μ M. Equal volumes of the PCR products for each SNP of a reaction group were combined. A total of 15 μ l of the PCR mix was treated with 2 units exonuclease I and 5 units shrimp alkaline phosphatase (New England Biolabs) at 37 $^{\circ}$ for 1 hr to remove dNTP and PCR primers, followed by 75 $^{\circ}$ for 15 min to inactivate the enzymes. The extension reaction was performed in a 10- μ l mix, containing 3 μ l treated PCR mix, 5 μ l SNaPshot multiplex ready reaction mix (Applied Biosystems), 0.2 pmole of each extension primer, and Milli-Q water. The reaction mix was incubated in a GeneAmp 2400 thermal cycler (Applied Biosystems) for 25 cycles of 96 $^{\circ}$ for 10 sec, 50–56 $^{\circ}$ for 5 sec, and 60 $^{\circ}$ for 30 sec. Primer extension product was treated by 1 unit shrimp alkaline phosphatase at 37 $^{\circ}$ for 1 hr followed by 75 $^{\circ}$ for 15 min to prevent migration of unincorporated fluorescent ddNTP. SNaPshot products were analyzed on an ABI PRISM 3100 DNA sequencer with GeneScan software version 3.5 and GeneScan-120 size standard analysis parameter files (Applied Biosystems).

Phylogenetic analysis: A phylogenetic tree was constructed by the maximum parsimony method using PAUP version 4 (SWOFFORD 1998). Sequence base differences were manually input as a character matrix. Each indel was treated as one character regardless of the number of bases involved, and equal weight was given to indel and single-base substitution events. The clonal lineage of Typhimurium isolates and root of the phylogenetic tree was identified by comparison of sequences containing the polymorphisms observed in this study, with the homologous sequences of the five other *Salmonella enterica* serovars for which complete or near complete genome sequences are available: Serovar Typhi strains CT18 (PARKHILL *et al.* 2001) (EMBL accession no. AL513382) and Ty2 (DENG

et al. 2003) (GenBank accession no. AE014613), Serovar Paratyphi A (downloaded from <ftp://genome.wustl.edu/pub/seqmgr/bacterial/salmonella/S.paratyphiA>), Serovar Enteritidis and Serovar Pullorum (downloaded from <http://salmonella.utmem.edu>), and Serovar Gallinarum (downloaded from <ftp://ftp.sanger.ac.uk/pub/pathogens/Salmonella>).

RESULTS

Mutational changes in AFLP fragments: Four of the 18 AFLP fragments cloned and sequenced in our previous study are related to *S. enterica* chromosomal sequences (HU *et al.* 2002). They comprise two pairs of fragments, and in each pair one fragment corresponds to the sequence found in the LT2 genome and the other differs by mutation. We denote such AFLP fragments, which are affected by mutational changes, as mutation-related AFLP fragments. One of each pair of mutation-related AFLP fragments is likely to be present in the Typhimurium chromosome and, if located, both members of the pair could be amplified and sequenced using primers based on the available LT2 genome sequence to determine the nature of the differences in mutation-related AFLP fragments, as an alternative approach to cloning the fragments.

To search for other mutation-related AFLP fragments, we performed computer modeling of AFLP analysis of the LT2 genome for all 16 *MseI* +1/*EcoRI* +1 primer pair combinations. The 200 phylogenetically informative AFLP fragments observed in our previous study (HU *et al.* 2002) were compared with modeled LT2 AFLP fragments using selective base and fragment size. Twenty AFLP fragments, including the 2 previously cloned and sequenced fragments C/A-5 and C/C-1, were matched with modeled LT2 fragments. Segments of the chromosome that included the 18 newly located AFLP fragments were amplified and sequenced using LT2 genome-based primers (supplemental Table S1 at <http://www.genetics.org/supplemental/>). Amplicons of at least two isolates in which the identified AFLP fragment was either present or absent were sequenced (supplemental Table S5 at <http://www.genetics.org/supplemental/>).

Nature of mutation-related AFLP fragments: The sequences showed that these phylogenetically informative AFLP fragments resulted from mutational events (Table 1). Of the 20 AFLP fragments matched with modeled LT2 fragments, 15 were shown to be a member of an inversely correlated pair: only one fragment of each pair is present in any given isolate, and only one of each pair matched a modeled LT2 fragment. Four fragments fell into two correlated pairs, both fragments of each pair matching adjacent modeled LT2 fragments connected by an *EcoRI* restriction site, with both fragments absent in some isolates due to absence of the connecting *EcoRI* site. Another fragment was found to be a member of such a pair, in which the other was observed in the genome, but only inferred as an AFLP fragment as at

881 bp it was outside the AFLP size range analyzed. In total, 36 mutation-related AFLP fragments were studied including the two pairs, C/A-4 and C/A-5 and C/C-1 and C/C-2, previously cloned and sequenced (HU *et al.* 2002) and the inferred 881-bp fragment. The 32 new fragments were named by primer pair combination followed by fragment size in base pairs.

The differences between members of the 18 pairs of mutation-related AFLP fragments could be classified into four categories comprising 10 single-base substitutions and eight indels (Table 1). In four pairs, the fragments differed by a single-base substitution in an *EcoRI* or *MseI* restriction site: absence of an *EcoRI* site results in absence of two connected AFLP fragments; gain or loss of an *MseI* site generates a shorter or longer AFLP fragment. In two pairs, the fragments differed by a single-base substitution that affected an AFLP selective base and shifted the AFLP fragment into another primer pair combination. In four pairs, the fragments differed by a single-base substitution in an AFLP fragment that changed its mobility. In eight pairs, the fragments differed by an indel event, of which two led to gain or loss of an *EcoRI* site and six affected mobility of the AFLP fragment. Of the 10 single-base substitutions, 1 was in an IG and 9 in genes comprising three synonymous and six nonsynonymous substitutions. Of the eight indels, four caused frameshifts in genes, three in IGs, and one comprises both coding sequence and IG. The distribution of mutational changes in the 46 isolates studied is given in supplemental Table S5 (<http://www.genetics.org/supplemental/>).

Mutational changes in intergenic regions for phylogeny:

We needed more mutational changes in addition to those detected in AFLP fragments, to increase the reliability of the phylogenetic analysis, and looked for mutational changes in IGs. Mutational changes in IGs are thought to be generally neutral and have little or no selective cost, and the effective mutation rates for IGs, including likelihood of fixation, would be expected to be higher than those for housekeeping genes, where there are major constraints on mutation.

In the mutation-related AFLP fragments discussed above, the ratio of mutations in coding sequences and IGs is 3:1 (13.5:4.5 as 13 inside genes, four in IGs, and one in both as 0.5 each). Coding regions in the LT2 chromosome (excluding RNA and pseudogenes) comprise 4218 kb, RNA and pseudogenes comprise 72.7 kb, and IGs comprise 566.7 kb. If the mutation rates were the same in the coding regions and IGs, the ratio of the mutations would be 7.4:1 (4218 kb:566.7 kb). This suggested that IGs do have higher effective mutation rates than coding regions.

We also compared coding regions and IGs of Typhimurium LT2 and *S. enterica* serovar Typhi (denoted Typhi in this article) CT18 on the basis of their genome sequences. The sequences of 77 Typhi CT18 central intermediary metabolism genes (http://www.sanger.ac.uk/Projects/S_typhi/St_gene_list_hierarchical.shtml)

TABLE 1
Nature of mutational changes detected in mutation-related AFLP fragments

Mutation-related AFLP fragments	Matched Typhimurium LT2 sequence	Mutational changes ^a	Location in gene/IG	LT2 position ^b
Point mutations that affect <i>EcoRI</i> or <i>MseI</i> restriction site				
T/C-142 and A/A-227	4586313–4586427 and 4586427–4586628	A to G (s)	<i>yjeA</i>	4586427 (E ^c)
A/A-327 and T/A-590	3237412–3237713 and 3237412–3237976	T to G (ns)	<i>tktA</i>	3237711 (M ^c)
G/A-364 and G/A-271	2391801–2392139 and 2391801–2392044	G to A (ns)	<i>glpB</i>	2392046 (M ^c)
T/G-80 and T/G-48	3829221–3829275 and 3829221–3829241	G to A	IG (<i>yhjY</i> , <i>tag</i>)	3829243 (M ^c)
Point mutations that affect AFLP selective base				
A/C-182 and A/T-182	1007389–1007545	C to T (s)	STM0932	1007393
T/G-556 and T/A-556	4799034–4799564	G to A (ns)	STM4540	4799038
Point mutations inside AFLP fragments				
A/A-546 and A/A-546a	543965–544483	G to A (ns)	<i>htpG</i>	544160
G/G-169 and G/G-169a	4603878–4604019	T to A (s)	<i>miaA</i>	4603961
T/A-223 and T/A223a	3654169–3654366	G to A (ns)	<i>yrfI</i>	3654053
T/G-247 and T/G-247a	218082–218303	C to A (ns)	<i>ligT</i>	218107
Insertions and deletions (indels)				
T/C-881 (predict) and G/A-346	3061642–3062495 and 3062495–3062815	183 bp (d)	<i>ygbI</i> and IG (<i>ygbI</i> , STM2920)	3062413–3062595 (E ^c)
T/T-232 and T/T-231	791360–791566	T (d, f)	STM0725	791460
A/G-150 and A/G-149	456303–456427	G (d, f)	<i>malZ</i>	456364
T/C-263 and T/C-264	720452–720689	G (i, f)	<i>ybeU</i>	720484–720485
T/G-396 and T/G-377	481388–481758	19 bp (d, f)	<i>phnT</i>	481632–481650
A/T-109 and A/A-342	906954–907035 and 907035–907351	size not known ^d (d)	IG (<i>ybiT</i> , STM0839)	907035 ± (E ^c)
C/A-4 and C/A-5	1096650–1096833	4 bp (dp)	IG (STM1003, <i>pncB</i>)	1096771–1096774
C/C-1 and C/C-2	1766735–1766881	8 bp (dp)	IG (STM1672, STM1673)	1766792–1766799

^as, synonymous substitution; ns, nonsynonymous substitution; d, deletion; i, insertion; f, frameshift mutation; dp, duplication. Mutation forms of AFLP fragment sequence relative to LT2 have been deposited with the GenBank Data Libraries under accession nos. DQ077505 to DQ077519.

^bPosition numbers are according to LT2 chromosome sequence (download from <http://genome.wustl.edu/gsc/bacterial/salmonella.shtml>).

^cE, affect *EcoRI* site; M, affect *MseI* site.

^dFragments A/T-109 and A/A-342 could not be amplified in DT64 using primers based on LT2 sequence up to 2 kb away, but can be amplified in all other DTs studied. An *EcoRI* site connecting these two fragments was lost due to a deletion event confirmed by Southern hybridization (data not shown).

were compared with those of Typhimurium LT2. The pairwise percentage differences for each pair ranged from 0 to 4.45% with an average of 1.20% (supplemental Table S6 at <http://www.genetics.org/supplemental/>). The IGs of the LT2 and Typhi CT18 chromosomes were obtained from the genome sequences and named sequentially TM1 or TY1, etc. BlastN searches gave matched pairs, which were confirmed by sequence alignment. The pairwise differences for each pair ranged from 0 to 21.30% with an average of 3.12% (data not shown). This confirms that IGs have higher effective mutation rates than housekeeping genes as expected.

We selected 48 IGs with relatively large size and/or high pairwise differences between LT2 and Typhi CT18,

ranging from 1.93 to 21.30% with an average of 5.13% (supplemental Table S7 at <http://www.genetics.org/supplemental/>). The selected IGs were amplified and sequenced using LT2 genome-based primers (supplemental Table S2 at <http://www.genetics.org/supplemental/>) for one isolate of each phage type (supplemental Table S8 at <http://www.genetics.org/supplemental/>). The primers were in the flanking genes, and that part of the genes for which we obtained good sequence was included in the analysis.

The sequences revealed 33 polymorphic sites in 23 IG/flanking gene regions. One was presence or absence of IS200 (TM781, between LT2 bases 874139 and 874140), which has six copies in LT2 (LAM and ROTH

TABLE 2
Details of single nucleotide polymorphisms in
IG / flanking gene regions

SNP name ^a	SNP detected ^b	LT2 position ^c
TM81 ^d	G to A (ns, STM0080)	93754
TM110_1	C to A	128947
TM110_2	T to C	128994
TM594 ^d	C to T (s, <i>dsbG</i>)	669268
TM716	G to A	801325
TM970_1	G to T	1091061
TM970_2	G to A	1091092
TM970_3	A to G	1091323
TM973 ^d	C to T (s, STM1002)	1095018
TM1231 ^d	G to T (ns, STM1269)	1350282
TM1809_1	G to A	1955487
TM1809_2 ^d	G to A (nm, STM1858)	1955831
TM1883	T to C	2026918
TM1899	T (d ^e)	2040507
TM2076	G to A	2231167
TM2079	A to T	2233518
TM2140_1	A (i ^e)	2298577
TM2140_2	A to G	2298714
TM2448	T to G	2650938
TM2461 ^d	C to T (ns, <i>pbpC</i>)	2665641
TM2488	C to A	2698341
TM2702	G to A	2928257
TM3002	C to G	3252889
TM3124_1	G to T	3380076
TM3124_2	A to T	3380089
TM3211_1	T to C	3474470
TM3211_2 ^d	A to G (s, <i>nlp</i>)	3474609
TM3230_1	T to C	3487935
TM3230_2	C to A	3488112
TM3230_3 ^d	G to T (s, <i>yrbL</i>)	3488194
TM3275_1 ^d	G to A (ns, <i>yhdP</i>)	3539590
TM3275_2	G to C	3539842

^a SNP name is the IG name where SNPs found. Where there is more than one SNP in an IG, the name includes a number following the IG name.

^b Mutation forms of IG / flanking gene regions sequence relative to LT2 have been deposited with the GenBank Data Libraries under accession nos. DQ077520 to DQ077552.

^c Position numbers are according to LT2 chromosome sequence (download from <http://genome.wustl.edu/gsc/bacterial/salmonella.shtml>).

^d SNP in flanking gene. s, synonymous substitution; ns, non-synonymous substitution; nm, nonsense mutation.

^e d, deletion; i, insertion.

1983), and 32 were SNPs, of which 30 were single-base substitutions and 2 were single-base indels. Nine of the 30 substitutions were in flanking genes, comprising four synonymous and four nonsynonymous substitutions, and one nonsense mutation. The details of SNPs are given in Table 2.

The distribution of mutational changes in the remaining 37 isolates was then determined. For the presence or absence of IS200, PCR was performed and scored visually after electrophoresis. Of the 32 SNPs, 14

were investigated further by PCR-RFLP as the mutation affected a restriction site, 14 by SNaPshot, and the other 4 by sequencing. The distribution of SNPs and the IS200 insertion in the 46 isolates studied is given in supplemental Table S8 (<http://www.genetics.org/supplemental/>).

DISCUSSION

Phylogenetic relationships and evolution of Typhimurium: We combined the data on mutational changes discussed above and obtained a phylogenetic tree on the basis of parsimony (Figure 1). Use of homologous sequences in the genome sequences for *S. enterica* serovars Typhi, Paratyphi A, Gallinarum, Enteritidis, and Pullorum enabled us to infer the direction for 50 of the 51 mutations, as one form of the polymorphism in Typhimurium isolates is found in all or most of the five other serovars and is assumed to be the ancestral genotype. For one locus (TM3230_3), a third form was present in the five other serovars, so it is not possible to determine the ancestral genotype and, as it is the only event on branch 10, the direction can be inferred only from that of the adjacent branches. In all branches with multiple events, the inferred direction of change was the same, and the 50 mutations gave a consistent pattern with the inferred root of the tree indicated in Figure 1. This consistency allows us to identify the putative ancestral genotype and gives us confidence in the clonal lineage of the Typhimurium isolates studied. As we scored only the base(s) that varied within Typhimurium, the other five serovars were not included in the tree.

The mutation-based tree grouped most of the Typhimurium isolates according to phage type, making it possible to give a single origin on the tree for all isolates of DT9, DT12a, DT64, DT108, DT126, DT135, and DT141, although the sample size for each DT is small. The DT12a isolates are assumed to have a single origin with the root of the tree being DT12a. The tree confirmed the hypothesis that DT64 arose from DT9 (MMOLAWA *et al.* 2002). Phage types DT1 and DT44 were found to have two independent origins even for the limited set of strains used. DT1 isolate M1879 has the same origin as five DT44 isolates, while DT44 isolate M1875 is well separated from the other five DT44 isolates, and clustered with three DT12a isolates (M1864–M1866) and two DT1 isolates (M1877 and M1880).

Only two parallel events were observed. One is the insertion of IS200, which affected branches 3, 14, and 16 of the tree. The other is a single-base substitution G to T (TM3124_1) in an IG, which is found in branches 6 and 11. The above events are shown in Figure 1. The distribution of each of the remaining 49 polymorphisms, which are all independent, can be explained by a single event. Even when two or three polymorphic sites are in the same IG, the mutations occurred in different branches, with no evidence of recombination involving two or

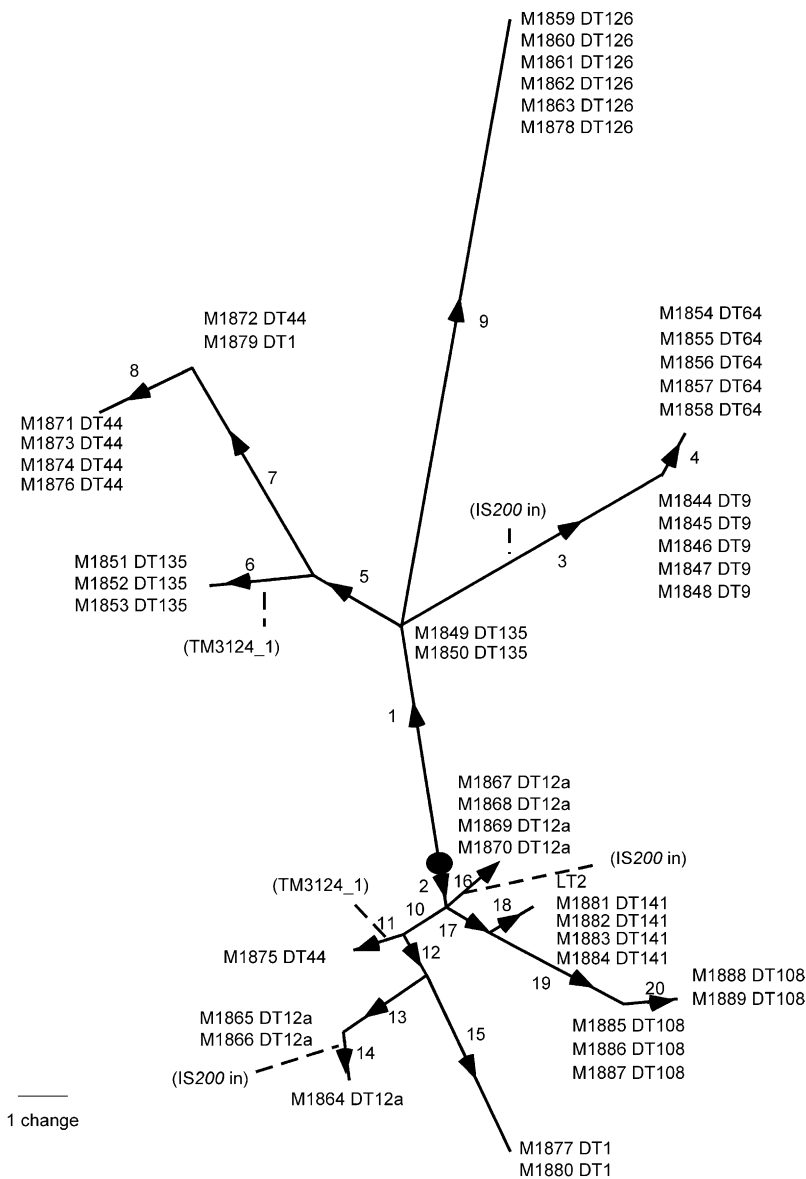


FIGURE 1.—Phylogenetic tree based on mutational changes by the maximum parsimony method using PAUP. The root of the tree is indicated by a solid circle. The direction of evolution in the Typhimurium isolates is indicated by arrows on the branches. The branch lengths are proportional to the number of events. The details of events that occurred on each branch (1–20) are: (1) A to G (TM970_3), T to C (TM3230_1), T to A (G/G-169 and G/G-169a), insertion of G (T/C-263 and T/C-264), G to A (T/A-556 and T/G-556); (2) insertion of A (TM2140_1); (3) C to A (TM110_1), A to G (TM2140_2), T to G (TM2448), 4-bp duplication (C/A-4 and C/A-5), 19-bp deletion (T/G-377 and T/G-396), IS200 insertion (TM781); (4) deletion (A/T-109 and A/A-342); (5) G to A (TM1809_2), T to C (TM1883); (6) G to T (TM3124_1), T to G (A/A-327 and T/A-590); (7) G to A (TM970_2), A to G (TM3211_2), C to A (TM3230_2), G to A (T/A-223 and T/A-223a), C to A (T/G-247 and T/G-247a); (8) G to A (TM716), C to T (TM2461); (9) T to C (TM110_2), C to T (TM594), G to T (TM970_1), G to A (TM1809_1), G to A (TM2702), C to G (TM3002), T to C (TM3211_1), G to A (TM3275_1), C to T (A/C-182 and A/T-182), 8-dp (C/C-1 and C/C-2), deletion of T (T/T-231 and T/T-232), deletion of G (A/G-149 and A/G-150), 183-bp deletion (G/A-346 and T/C-881); (10) G to T (TM3230_3); (11) G to T (TM3124_1); (12) A to T (TM2079); (13) A to T (TM3124_2), G to A (G/A-364 and G/A-271); (14) IS200 insertion (TM781); (15) G to A (A/A-546 and A/A-546a), A to G (A/A-227 and T/C-142), G to A (T/G-48 and T/G-80), deletion of T (TM1899); (16) IS200 insertion (TM781); (17) C to T (TM 973); (18) C to A (TM2488); (19) G to A (TM81), G to A (TM2076), G to C (TM3275_2); (20) G to T (TM1231). Note that IS200 insertion (TM781) occurred in branches 3, 14, and 16. G to T (TM3124_1) occurred in branches 6 and 11. G to T (TM3230_3) in branch 10 could not be inferred from the other genomes as a third form was found in all other five serovars, but can be inferred from the direction of change in adjacent branches.

more sites. This indicates that Typhimurium has a very strong clonal structure with a very low level of recombination over the time for its diversification.

High proportion of nonsynonymous mutations:

Interestingly, we found more nonsynonymous than synonymous substitutions for those in coding regions. There are 10 nonsynonymous and 7 synonymous substitutions, one nonsense mutation, and four frameshift mutations. The high ratio of nonsynonymous substitutions is consistent with our previous studies on the *gnd* locus of *S. enterica* (THAMPAPILLAI *et al.* 1994) and *Escherichia coli* (BISERCIC *et al.* 1991), where a higher ratio of nonsynonymous substitutions was found within each species than between them (THAMPAPILLAI *et al.* 1994), and is attributed to the threshold for slightly deleterious mutations to be fixed within some clones of a species being less stringent than for fixation over the larger population and longer time in a species. The

effect is more marked in this study, which is to be expected as Typhimurium is one clone of one of the two species used in the previous study and hence has a smaller population and even lower level of stringency. A more traditional explanation of high levels of nonsynonymous substitutions is that the genes concerned are changing under selection pressure. We consider this unlikely in this case as 7/10 of nonsynonymous substitutions are in genes that are found in nine genomes of three related species (*E. coli* K-12, *E. coli* O157:H7, *Klebsiella pneumoniae* and six *S. enterica* serovars) used in an analysis of gene distributions (McCLELLAND *et al.* 2001), and it seems unlikely that such conserved genes would be under selection pressure for change. Moreover, a similar finding has been reported by FEIL *et al.* (2003) in a study of *Staphylococcus aureus*. It may generally be found that there is a higher proportion of nonsynonymous substitutions in comparisons of closely related

isolates than in comparisons at species or higher taxon level, and the phenomenon is worthy of further study.

Advantages of sequencing IGs in detecting polymorphisms: The level of polymorphism in IGs is higher than that in housekeeping genes, the average pairwise difference between LT2 and CT18 sequences being 1.20% for central intermediary metabolism genes and 5.13% for the 48 IGs studied, a factor of approximately four. The average length of the 48 IGs studied was 300 bp, so that a single sequence run (~600 bp) would be able to cover one IG and part of adjacent genes. If divergence was such that a single sequence run of a housekeeping gene could detect two mutations, it would on average detect five polymorphisms by sequencing an IG and adjacent coding DNA (four in the IG and one in adjacent coding DNA). Furthermore, mutational changes in IGs are thought to be generally neutral, offering further advantages in a molecular evolution study.

Comparison of relationships of Typhimurium isolates inferred from mutational changes with that derived from AFLP data: It is interesting to compare the phylogenetic relationships of Typhimurium isolates inferred from mutational changes (Figure 1, mutation-based tree) with the dendrograms derived previously from AFLP fingerprinting patterns (AFLP tree) (HU *et al.* 2002). Since AFLP picks up both DNA gain or loss polymorphisms and mutational changes, the major differences between the two trees are due to inclusion of DNA gain or loss polymorphisms. We are now able to relate the gain or loss of the informative AFLP fragments observed (HU *et al.* 2002) to specific nodes in the mutation-based tree. Three examples are discussed in the following paragraphs.

DT9 and DT64 isolates are closely related in both mutation-based and AFLP trees, but branch lengths differ. In the mutation-based tree, the DT64 isolates are seen to derive from DT9, with only a single mutation distinguishing them. AFLP analysis revealed six fragments in DT64 only, three of which (C/A-1, C/G-1, and G/A-1) have been sequenced and are all lambdaoid phage gene related (HU *et al.* 2002). The data indicated that DT9 and DT64 are closely related and that DT64 was derived from DT9 by gain of phage-related DNA. Our results are consistent with the report by MMOLAWA *et al.* (2002) that Typhimurium DT9 can be converted to DT64 by a phage induced from DT64.

For our second example, we look at DT141 and DT108, which are closely related in the mutation-based tree, but well separated in the AFLP tree. AFLP analysis revealed 12 fragments in DT108 only, and 5 fragments in DT12a and DT108 only, 2 in DT135 and DT108 only, and 2 in DT64, DT12a, and DT108 only. Four of these fragments (C/T-1, C/A-2, C/A-3 and G/C-1) have been sequenced. C/T-1 is phage related, C/A-2 and C/A-3 are due to gain of DNA of unknown origin (HU *et al.* 2002), and G/C-1 has 98% similarity to *Shigella flexneri* phage Sf6 gene 34 (CASJENS *et al.* 2004). The data indicated

that DT108 and DT141 have the same origin, but DT108 gained extra DNA, probably phage related, some of which was gained independently in DT12a or DT64, which affected the relationships in the AFLP tree.

Our third example is the distribution of DT1 isolates in two divergent parts of the mutation-based tree. DT1 isolates M1877 and M1880 are related to three DT12a isolates, whereas, DT1 isolate M1879 is related to five DT44 isolates on a different major branch of the mutation-based tree and is presumably of independent origin. These three DT1 isolates are closely related in the AFLP tree. AFLP analysis revealed 10 fragments specific to DT1. Three of them (C/G-2, C/T-2, and C/A-7) have been sequenced and are all related to plasmid R64 genes (HU *et al.* 2002). Independent acquisition of the same plasmid on two occasions probably determined the phage type of the three DT1 isolates, and the presence of 10 fragments common to all DT1 isolates pulled them together in the AFLP tree.

Conclusions: A phylogenetic tree of *S. enterica* serovar Typhimurium based on mutational changes gave a clear picture of the relationships of the 46 isolates studied, which had been obscured in several places in the previously reported AFLP tree by the more common fragments from mobile genetic elements. Comparison with genomes of other serovars identified the ancestral form for almost all polymorphisms, gave a root for the tree and a clear clonal lineage of the Typhimurium isolates studied. The high level of consistency in the data gives confidence in the tree and allows one to relate other events to the tree. The mutation-based tree also indicated a very low level of recombination during the diversification of Typhimurium as the clonal variations observed are from point mutations and, with exception of IS200 insertion and TM3124_1, none are attributed to recombination.

We found that most Typhimurium isolates of a given phage type are in the same evolutionary group, but also that some phage types appear to have arisen more than once. Comparison of the phylogenetic tree with AFLP data gave examples of unrelated isolates of a given phage type having common AFLP fragments comprising plasmid or phage genes, supporting the view that phage type can be determined by presence of specific phages or plasmids.

This study provides the first detailed analysis of relationships of isolates within a clone. The level of variation was far too low for MLST to be efficient (FAKHR *et al.* 2005), and sequencing IGs provided a good strategy for identifying mutational polymorphisms and for studying the molecular evolutionary relationships of closely related isolates. The methods used to determine distribution of the alleles, being based on sequence, are easily extended to include much wider representation of Typhimurium and allow for deeper analysis of the diversification of this very successful clone. The establishment of a reliable phylogeny for isolates within

Typhimurium provides a basis for studying the adaptation of various forms of this serovar, which has been shown to have host-adapted variants (BÄUMLER *et al.* 1998; RABSCH *et al.* 2002). Given the relatively low level of variation within even a quite diversified clone such as Typhimurium, it may be much easier to identify the defining events in matters such as host adaptation than in comparisons of the more distantly related organisms often studied, where there is likely to be more variation not related to the adaptations of interest.

We are grateful to Chris Murray of the Institute of Medical and Veterinary Science (IMVS), Adelaide, South Australia, Australia, for kindly providing strains. We thank Bing Yu in the Department of Molecular and Clinical Genetics, Royal Prince Alfred Hospital (Sydney), for advice on SNaPshot. This research was supported by a grant from the National Health and Medical Research Council of Australia.

LITERATURE CITED

- ALLAND, D., T. S. WHITTAM, M. B. MURRAY, M. D. CAVE, M. H. HAZBON *et al.*, 2003 Modeling bacterial evolution with comparative-genome-based marker systems: application to mycobacterium tuberculosis evolution and pathogenesis. *J. Bacteriol.* **185**: 3392–3399.
- ANDERSON, E. S., L. R. WARD, M. J. SAXE and J. D. DE SA, 1977 Bacteriophage typing designations of *Salmonella typhimurium*. *J. Hyg.* **78**: 297–300.
- BÄUMLER, A. J., R. M. TSOLIS, T. A. FICHT and L. G. ADAMS, 1998 Evolution of host adaptation in *Salmonella enterica*. *Infect. Immun.* **66**: 4579–4587.
- BISERCIC, M., J. Y. FEUTRIER and P. R. REEVES, 1991 Nucleotide sequence of the *gnd* gene from nine natural isolates of *Escherichia coli*: evidence of intragenic recombination as a contributing factor in the evolution of the polymorphic *gnd* locus. *J. Bacteriol.* **173**: 3894–3900.
- CASJENS, S., D. A. WINN-STAPLEY, E. B. GILCREASE, R. MORONA, C. KUHLEWEIN *et al.*, 2004 The chromosome of shigella flexneri bacteriophage Sf6: complete nucleotide sequence, genetic mosaicism, and DNA packaging. *J. Mol. Biol.* **339**: 379–394.
- DENG, W., S.-R. LIU, G. PLUNKETT, III, G. F. MAYHEW, D. J. ROSE *et al.*, 2003 Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18. *J. Bacteriol.* **185**: 2330–2337.
- FAKHR, M. K., L. K. NOLAN and C. M. LOGUE, 2005 Multilocus sequencing typing lacks the discriminatory ability of pulsed-field gel electrophoresis for typing *Salmonella enterica* serovar Typhimurium. *J. Clin. Microbiol.* **43**: 2215–2219.
- FEIL, E. J., J. E. COOPER, H. GRUNDMANN, D. A. ROBINSON, M. C. ENRIGHT *et al.*, 2003 How clonal is *Staphylococcus aureus*? *J. Bacteriol.* **185**: 3307–3316.
- GLYNN, M. K., C. BOPP, W. DEWITT, P. DABNEY, M. MOKHTAR *et al.*, 1998 Emergence of multidrug-resistant *Salmonella enterica* serotype typhimurium DT104 infections in the United States. *N. Engl. J. Med.* **338**: 1333–1338.
- GORDON, D., C. ABAJIAN and P. GREEN, 1998 CONSED—A graphical tool for sequence finishing. *Genome Res.* **8**: 195–202.
- GUTACKER, M. M., J. C. SMOOT, C. A. L. MIGLIACCIO, S. M. RICKLEFS, S. HUA *et al.*, 2002 Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms: resolution of genetic relationships among closely related microbial strains. *Genetics* **162**: 1533–1543.
- HU, H., R. LAN and P. R. REEVES, 2002 Fluorescent amplified-fragment length polymorphism analysis of *Salmonella enterica* serovar Typhimurium reveals phage-type specific markers and potential for microarray typing. *J. Clin. Microbiol.* **40**: 3406–3415.
- KOTETISHVILI, M., O. C. STINE, A. KREGER, J. G. J. MORRIS and A. SULAKVELIDZE, 2002 Multilocus sequence typing for characterization of clinical and environmental *Salmonella* strains. *J. Clin. Microbiol.* **40**: 1626–1635.
- LAM, S., and J. R. ROTH, 1983 IS200: a *Salmonella*-specific insertion sequence. *Cell* **34**: 951–960.
- LOW, J. C., M. ANGUS, G. HOPKINS, D. MUNRO and S. C. RANKIN, 1997 Antimicrobial resistance of *Salmonella enterica* typhimurium DT104 isolates and investigation of strains with transferable apramycin resistance. *Epidemiol. Infect.* **118**: 97–103.
- MCCLELLAND, M., K. E. SANDERSON, J. SPIETH, S. W. CLIFTON, P. LATREILLE *et al.*, 2001 Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* **413**: 852–856.
- MMOLAWA, P. T., R. WILLMORE, C. J. THOMAS and M. W. HEUZENROEDER, 2002 Temperate phages in *Salmonella enterica* serovar Typhimurium: implications for epidemiology. *Int. J. Med. Microbiol.* **291**: 633–644.
- PARKHILL, J., G. DOUGAN, K. D. JAMES, N. R. THOMSON, D. PICKARD *et al.*, 2001 Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* **413**: 848–852.
- POWLING, J. (Editor), 1997a *National Enteric Pathogens Surveillance Scheme: Human Annual Report 1996*. Microbiological Diagnostic Unit, The University of Melbourne, Victoria, Australia.
- POWLING, J. (Editor), 1997b *National Enteric Pathogens Surveillance Scheme: Non-Human Annual Report 1996*. Microbiological Diagnostic Unit, The University of Melbourne, Victoria, Australia.
- RABSCH, W., H. L. ANDREWS, R. A. KINGSLEY, R. PRAGER, H. TSCHAPE *et al.*, 2002 *Salmonella enterica* serotype Typhimurium and its host-adapted variants. *Infect. Immun.* **70**: 2249–2255.
- READ, T. D., S. L. SALZBERG, M. POP, M. SHUMWAY, L. UYAMAY *et al.*, 2002 Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science* **296**: 2028–2033.
- REEVES, P. R., L. FARNELL and R. LAN, 1994 MULTICOMP: a program for preparing sequence data for phylogenetic analysis. *Comput. Appl. Biosci.* **10**: 281–284.
- STADEN, R., 1994 Staden: searching for restriction sites. *Methods Mol. Biol.* **25**: 79–85.
- SWOFFORD, D. L., 1998 *PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods)*. Sinauer Associates, Sunderland, MA.
- THAMPAPILLAI, G., R. LAN and P. R. REEVES, 1994 Molecular evolution in the *gnd* locus of *Salmonella enterica*. *Mol. Biol. Evol.* **11**: 813–828.
- THOMPSON, J. D., D. G. HIGGINS and T. J. GIBSON, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- VOS, P., R. HOGERS, M. BLEEKER, M. REIJANS, T. VAN DE LEE *et al.*, 1995 AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**: 4407–4414.

Communicating editor: N. TAKAHATA