Characterization of the Genomes of a Diverse Collection of *Salmonella enterica* Serovar Typhimurium Definitive Phage Type 104[∇]

Fiona J. Cooke,[1]* Derek J. Brown,[2] Maria Fookes,[1] Derek Pickard,[1] Alasdair Ivens,[1] John Wain,[1] Mark Roberts,[3] Robert A. Kingsley,[1] Nicholas R. Thomson,[1] and Gordon Dougan[1]

Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom[1]; Scottish Salmonella Reference Laboratory, Stobhill Hospital, 133 Balornock Road, Glasgow G21 3UW, Scotland[2]; and Faculty of Veterinary Medicine, 464 Bearsden Road, Glasgow G61 1QH, Scotland[3]

Salmonella enterica serovar Typhimurium definitive phage type 104 (DT104) has caused significant morbidity and mortality in humans and animals for almost three decades. We completed the full DNA sequence of one DT104 strain, NCTC13348, and showed that significant differences between the genome of this isolate and the genome of the previously sequenced strain *Salmonella* serovar Typhimurium LT2 are due to integrated prophage elements and *Salmonella* genomic island 1 encoding antibiotic resistance genes. Thirteen isolates of *Salmonella* serovar Typhimurium DT104 with different pulsed-field gel electrophoresis (PFGE) profiles were analyzed by using multilocus sequence typing (MLST), plasmid profiling, hybridization to a pan-*Salmonella* DNA microarray, and prophage-based multiplex PCR. All the isolates belonged to a single MLST type, sequence type ST19. Microarray data demonstrated that the gene contents of the 13 DT104 isolates were remarkably conserved. The PFGE DNA fragment size differences in these isolates could be explained to a great extent by differences in the prophage and plasmid contents. Thus, here the nature of variation in different *Salmonella* serovar Typhimurium DT104 isolates is further defined at the gene and whole-genome levels, illustrating how this phage type evolves over time.

Salmonella enterica is a globally important cause of infections in economically important animal groups, as well as humans. *S. enterica* isolates are classified on the basis of DNA homology and phenotypic characteristics into subspecies, and subspecies I is the most common cause of infection in humans and domestic animals (3). *S. enterica* subspecies I is routinely further subclassified into serovars using the Kauffman-White scheme, based on antigenic variation in O lipopolysaccharide and H flagellum antigens (27). Over the past century this scheme has proved to be extremely valuable in supporting the epidemiology and clinical diagnosis of salmonellosis. We know that certain strains belonging to serovars of *S. enterica*, often classified by phage typing, cause epidemic outbreaks, whereas other strains are associated with sporadic or epidemic infections. Further, epidemic strains may be dominant in a particular geographical region for a period of time and may eventually be replaced by a distinct strain. *S. enterica* strains belonging to particular serovars are normally identified by a combination of phenotypic and genotypic approaches. In addition to phage typing, phenotyping approaches can involve biotyping and antibiotic resistance profiling, whereas genotyping approaches include pulsed-field gel electrophoresis (PFGE) and PCR-based assays. These approaches can distinguish between strains, but they provide limited information about the genetic differences between isolates and do not readily determine how isolates classified in an epidemic group evolve during an epidemic. Genome-wide approaches may be able to provide such information.

Phage typing, which is often an initial step in subclassifying *S. enterica* serovar Typhimurium strains, showed that definitive phage type 104 (DT104) isolates began to emerge as a problem in the 1980s and soon spread to humans and domestic animals around the world (33). DT104 isolates were frequently resistant to multiple antibiotics, which was associated with the acquisition of the mobile genetic element *Salmonella* genomic island 1 (SGI1) (5). At present, information about the diversity of DT104 isolates and about differences between DT104 isolates and other *Salmonella* serovar Typhimurium phage types is limited. The first *Salmonella* serovar Typhimurium genome to be fully sequenced and annotated was the genome of an isolate of strain LT2 (23). Now the fully annotated sequence of a multiply antibiotic-resistant DT104 strain, strain NCTC13348 isolated from a human case of gastroenteritis in 1988 (28), is available (http://www.sanger.ac.uk/Projects/Salmonella/). Here we exploited these *Salmonella* serovar Typhimurium genome sequences together with a collection of DT104 isolates with 13 different PFGE profiles to characterize the genomes of DT104 isolates. To this end, we employed a combination of bioinformatic and molecular approaches, including plasmid profiling, prophage multiplex PCR (7), DNA microarray analysis, and multilocus sequence typing (MLST) (10, 18, 21). The results of these studies are presented below.

MATERIALS AND METHODS

Bacterial strains. Thirteen *Salmonella* serovar Typhimurium DT104 isolates (isolates D1 to D13) were provided by The Scottish *Salmonella* Reference Lab-

* Corresponding author. Mailing address: Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom. Phone: 44 (0) 1223 495398. Fax: 44 (0) 1223494919. E-mail: fiona@sanger.ac.uk.

TABLE 1. *Salmonella* serovar Typhimurium DT104 isolates

Isolate	Year of isolation	Laboratory	R type[a]	Plasmid size(s) (kb)	PFGE profile
D1	2004	West Lothian	ACSSuSpT	90, 60, 50	STYMXB.0001 + 50 kb
D2	2004	Ayrshire	ACSSuSpTTm	90, 60, 11, 6.8, 5.7	STYMXB.0009
D3	2005	Aberdeen	ACSSuSpT	90	STYMXB.0001
D4	2004	Dundee	SSuSpT	95, 90, 3.3	STYMXB.0001 + 95 kb
D5	2005	Glasgow	ACSSuSpTTm	7.0, 5.7, 4.8, 3.3, 3.2, 2.0	STYMXB.0006
D6	2003	Dumfries	ACSSuSpTTm	7.2, 3.2, 2.0	STYMXB.0006-660
D7	2005	Aberdeen	ACSSuSpT	90	STYMXB.0031
D8	2003	Dumfries	ACSSuSpTTm	90, 7.2, 7.0, 3.2, 2.0	Tm104X18
D9	2003	Edinburgh	ACSSuSpT	90, 40	STYMXB.0001 + 40 kb
D10	2004	Aberdeen	ACSSuSpT	90	STYMXB.0081
D11	2004	Aberdeen	ACSSuSpT	90, 60	Tm104XSco5
D12	2004	Falkirk	ACSSuSpT	90	TmXSco9
D13	2001	Aberdeen	ACSSuSpT	90	STYMXB.0074

[a] Antibiotic abbreviations: A, ampicillin; C, chloramphenicol; S, streptomycin: Su, sulfamethoxazole; Sp, spectinomycin; T, tetracycline: Tm, trimethoprim.

oratory. These isolates were all isolated from human cases of gastroenteritis between 2001 and 2005 and were selected so that they represented different PFGE groups (32) (Table 1). The patients were from different regions of Scotland, and their ages ranged from 3 to 76 years. All isolates were confirmed to be *Salmonella* serovar Typhimurium DT104 by using culture and serology (National Standard Methods [http://www.hpa-standardmethods.org.uk/documents /bsopid/pdf/bsopid24.pdf]). Phage typing was performed using standard methods (4). Antibiotic susceptibility (R-type) was determined by breakpoint agar incorporation of antibiotics for 21 agents, including ampicillin, chloramphenicol, streptomycin, sulfamethoxazole, spectinomycin, tetracycline, nalidixic acid, ciprofloxacin, furazolidone, and trimethoprim (14).

A *Salmonella* serovar Typhimurium DT104 strain in which the *dam* gene was replaced (strain RAK102) with the aminoglycoside phosphotransferase gene (*aph*) encoding kanamycin resistance was constructed by allelic exchange by using a previously described method, with minor modifications (8). Primers RKS248 (5′-CAGAATTGAGGGGGCAATCAAATACTGTTTCATCCGCTT CTCCTTGAGAATGTGTAGGCTGGAGCTGCTTCG-3′) and RKS249 (5′-G CTGTCGGAGCTTTCTCCACAGCCGGAGAAGGTGTAATTAGTTAGT CAGCCATATGAATATCCTCCTTAG-3′) were used to amplify the *aph* gene from pKD4 (8). These primers contain a 50-nucleotide sequence at the 5′ end identical to the flanking sequence of the intended deletion (positions 1 to 837 of the *dam* open reading frame). The resulting PCR product was introduced by electrotransformation into *Salmonella* serovar Typhimurium strain SL3261 containing plasmid pKD46, in which the λ Red recombinase gene was induced with 10 mM L-arabinose. Recombination of the PCR product into the *Salmonella* serovar Typhimurium chromosome was selected by plating transformants on LB agar containing kanamycin. The *aph* gene was transferred to *Salmonella* serovar Typhimurium DT104 strain NCTC13348 by P22 transduction. One colony was selected, purified from P22 by streaking to obtain single colonies, and designated RAK102. PCRs using combinations of a forward oligonucleotide primer that annealed within the *aph* gene (5′-AGCTGTGCTCGACGTTGTCAC-3′) and forward primer 5′-CAGGCGAGCAAAATCAGCC-3′ and reverse primer 5′-G CACAGCCTAAAGCGCAGG-3′ flanking the *dam* deletion with DNA prepared from *Salmonella* serovar Typhimurium RAK102 (Δ*dam*::*aph*) was used to confirm the correct structure of the mutation (data not shown).

PFGE. PFGE was performed by using the Pulse-Net protocol with *Salmonella* serovar Braenderup H9812 as the marker (32). A statistical analysis was performed with BioNumerics software (Applied Maths, St-Martens-Latern, Belgium) with 1% tolerance and 0.8% optimization, using Dice coefficients to compare profiles. A dendrogram was constructed by using the unweighted-pair group method with arithmetic means. Fragments less than 30 kb long were not included in the final analysis (26). The STYMXB nomenclature of PFGE profiles is based on the SalmGene classification (26), now superseded by PulseNet Europe (http://www.cdc.gov/pulsenet/index.htm). Tm104X designations are specific to the Scottish database and were employed when there were no matches for a profile in the PulseNet database.

Plasmid profiling. Plasmids were identified using the method of Kado and Liu (17). Plasmid sizes were calculated by comparison with transconjugant *Escherichia coli* 39R861 (36).

MLST. MLST was performed by determining the sequences of seven housekeeping genes (*aroC*, *dnaN*, *hemD*, *hisD*, *purE*, *sucA*, and *thrA*) (18). The data

obtained were compared with the *Salmonella* MLST database at The Max Planck Institute (http://web.mpiib-berlin.mpg.de/mlst/dbs/Senterica).

Microarray analysis. Generation 3 of the PCR-product spotted *Salmonella* microarray constructed at the Wellcome Trust Sanger Institute has been described previously (7). The generation 3 microarray includes PCR products based on the genomes of *Salmonella* serovar Typhimurium strain LT2, DT104 strain NCTC13348, and strain SL1344, as well as open reading frames from the pSLT virulence plasmid. Genomic DNA extracted from *Salmonella* serovar Typhimurium DT104 isolates D1 to D13 was competitively hybridized to the *Salmonella* microarray using antibiotic-sensitive *Salmonella* serovar Typhimurium DT104 strain P247529 as a control (HPA, Colindale). Three slides were used for each isolate, with dye reversal. The washing procedures were stringent and included 200 ml of 2× SSC at room temperature for 5 min, two washes in 200 ml of 0.1× SSC–0.1% sodium dodecyl sulfate at 65°C with gentle agitation for 30 min, and two washes in 200 ml of 0.1× SSC at 65°C with gentle agitation for 30 min (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate). Slides were scanned using a Genepix 4000B scanner (Axon Instruments [now Molecular Devices], California), and every spot was assessed with Genepix Pro software (Axon Instruments). Data were normalized using GeneSpring software V7.2 (Silicon Genetics), and the final gene list consisted of microarray features present in all three slides for at least 1 of the 13 isolates studied. Features with a low signal (<200) in both raw and control channels for all isolates were ignored, thus reducing the final gene list to 4,110 features. Ratios with a raw signal of <200 in both test and control channels for only one isolate were treated with caution. Gene calling of hybridization ratios was performed by using the program GACK (19).

Phage multiplex PCR. A previously described multiplex PCR (7) was performed for isolates D1 to D13 using 15 oligonucleotide pairs based on the sequenced strain of *Salmonella* serovar Typhimurium DT104 (strain NCTC13348). These primer pairs targeted the five prophages and SGI1.

Microarray data accession number. The microarray raw data files have been deposited in the ArrayExpress database under accession no. E_MTAB-20.

RESULTS

Comparison of the genomes of *Salmonella* serovar Typhimurium DT104 strain NCTC13348 and strain LT2. We recently completed the genome sequence of *Salmonella* serovar Typhimurium DT104 strain NCTC13348 isolated from a human case of gastroenteritis that occurred in 1988 (http://www .sanger.ac.uk/Projects/Salmonella/). This sequence provided a framework with which to compare the genetic organization of diverse DT104 isolates. Initially, we compared the genome of *Salmonella* serovar Typhimurium DT104 strain NCTC13348 to that of *Salmonella* serovar Typhimurium LT2, which had been sequenced and annotated previously (23). Perhaps not surprisingly, the overwhelming feature of the results of this analysis
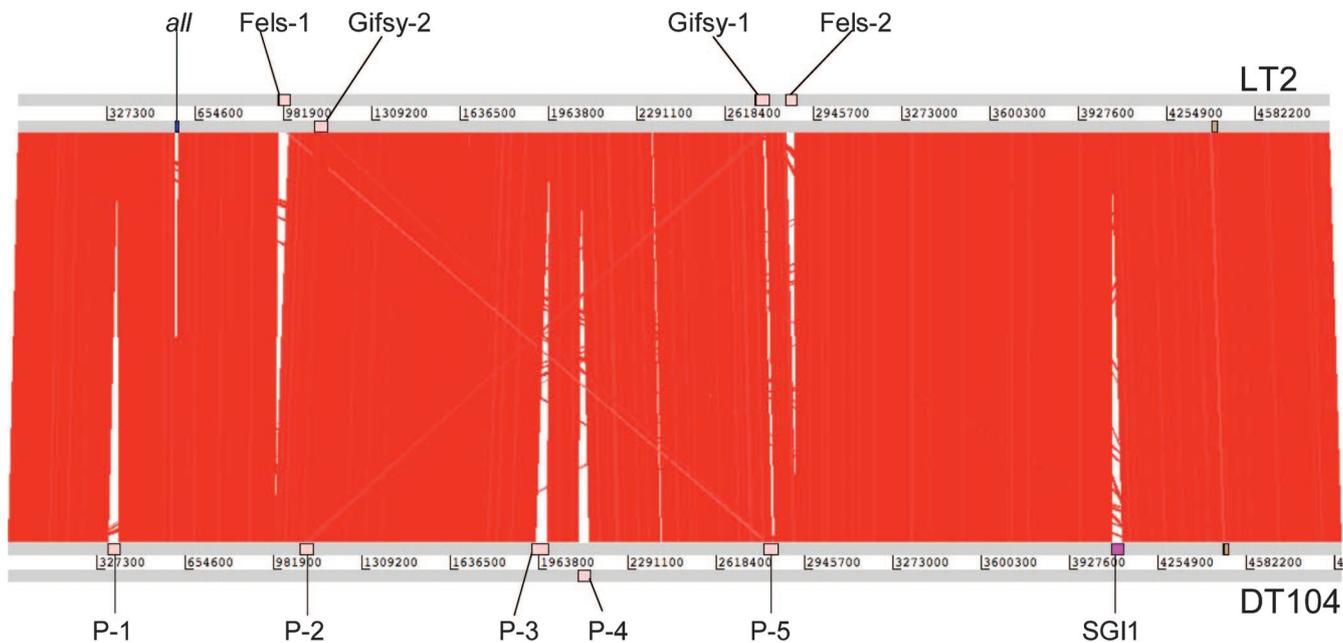
FIG. 1. ACT comparison of the genomes of LT2 and DT104 strain NCTC13348. DNA matches for the complete six-frame translations (computed using BLASTN) of the whole-genome sequences of *Salmonella* serovar Typhimurium LT2 (top) and *Salmonella* serovar Typhimurium DT104 strain NCTC113348 (bottom) were compared by using ACT (http://www.sanger.ac.uk/Software/ACT). Genome coordinates are indicated. The red bars between the DNA lines indicate individual BLASTN matches. DT104 prophages are indicated by arrows labeled P1 to P5, and the allantoin deletion is indicated by an arrow labeled *all*.

was the extremely high level of conservation of sequences, gene contents, and synteny between the two genomes. Figure 1 shows the results of visualization of this comparison using the Artemis Comparison Tool (ACT) (1, 6). At the level of genome resolution used, significant nonhomologous gene clusters could be readily detected, and most of the differences were associated with the mobile SGI1 and some prophagelike elements. When a more detailed analysis was performed, a number of smaller sequence differences were apparent, including the previously described loss of allantoin utilization genes (*all*) from the DT104 lineage (22). Consequently, we focused on the larger regions where there were differences for further analysis.

Salmonella serovar Typhimurium DT104 strain NCTC13348 harbors five prophage-related elements, designated P1 to P5 (Fig. 1), and a likely prophage remnant. Some of these elements are shared with *Salmonella* serovar Typhimurium LT2, including the complete GIFSY-2-like prophage 2 (genome coordinates 1079201 to 1124674), the left arm of prophage 5, which is related to the LT2 prophage GIFSY-1 (genome coordinates 2797168 to 2845750), and a prophage remnant (genome coordinates 4497317 to 4514057). Importantly, prophages GIFSY-1 and -2 have been shown to contribute to the virulence of *Salmonella* serovar Typhimurium in various infection models, including mice, cattle, and macrophage survival assays (11–13).

However, a number of phage-related elements were present in NCTC13348 but not present in LT2, including the apparently complete prophages 1 (genome coordinates 365588 to 406646), 3 (genome coordinates 1954176 to 1995367), and 4 (genome coordinates 2109720 to 2149064) and the central region of prophage 5 (genome coordinates 2797168 to 2845750).

Prophage 1 is a 41,058-bp P22-like phage (~70.4% homology with P22) previously described as PDT17 or ST104 (31). Prophage 3 is a 41-kbp phage that has homology with a phage from *Photorhabdus luminescens* (15). The open reading frame SDT1840 of DT104 strain NCTC13348 that precedes prophage 3 is also not present in LT2. Prophage 4 is a 39-kb phage that has some similarity with *Salmonella* serovar Typhimurium phage ST64B (24).

SGI1 has been described previously as a ~43-kb island that includes a 13-kb region harboring antibiotic resistance genes (5, 9, 25). Although SGI1 was initially regarded as unique to DT104, an SGI1-related element has been detected in other phage types of *Salmonella* serovar Typhimurium, in other *Salmonella* serovars, and recently in *Proteus mirabilis* (2). The SGI1 sequence in NCTC13348 was almost identical to the sequence determined for the previously sequenced SGI1 element except for a few single-nucleotide polymorphisms.

Genome variation of 13 *Salmonella* serovar Typhimurium DT104 isolates isolated between 2001 and 2005. Since we had access to the complete DNA sequence of a DT104 isolate, we were in a strong position to make genome-wide comparisons with *Salmonella* serovar Typhimurium DT104 field isolates collected at different times during a recent epidemic. We therefore analyzed DNA prepared from 13 such *Salmonella* serovar Typhimurium DT104 isolates that were collected at different times and places in Scotland and were known to have divergent PFGE patterns, which was indicative of genome variation. Initially, all 13 DT104 isolates were subjected to MLST analysis and were found to be indistinguishable and to belong to ST19 according to the Salmonella MLST database (http://web.mpiib-berlin.mpg.de/mlst/dbs/Senterica/). Figure 2 shows the PFGE profiles of the 13 *Salmonella* serovar Typhimurium
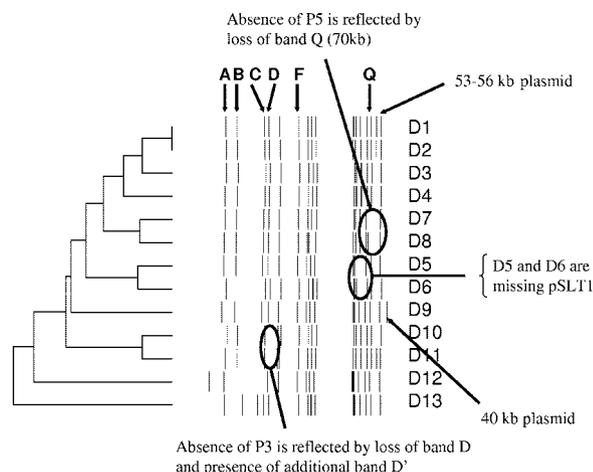
FIG. 2. Clustering of PFGE profiles generated by *Salmonella* serovar Typhimurium DT104 isolates D1 to D13. The dendrogram was constructed by the unweighted-pair group method with arithmetic means using BioNumerics software. Fragment sizes were obtained by comparison with *Salmonella* serovar Braenderup strain H9812 fragments, and fragments less than 30 kb long were not included in the final analysis. All fragments are not perfectly aligned as the image is a composite gel image.

DT104 isolates after cleavage with XbaI. Clear differences between the strains were observed, but analysis of the relationship between the patterns using BioNumerics software showed that the majority of them had Dice similarity coefficients greater than 90%, indicating that they were nevertheless highly related.

Isolates D1 to D13, representing three different antibiotic resistance types, ACSSuSpT ($n = 8$), ACSSuSpTTm ($n = 4$), and SSuSpT ($n = 1$) (where A indicates ampicillin, C indicates chloramphenicol, S indicates streptomycin, Su indicates sulfamethoxazole, Sp indicates spectinomycin, T indicates tetracycline, and Tm indicates trimethoprim), generated nine distinct plasmid profiles with plasmid sizes ranging from 95 to 2 kb. Five isolates (D3, D7, D10, D12, and D13) harbored only the 90-kb virulence-associated plasmid pSLT (16), whereas iso-

lates D5 and D6 harbored only plasmids that were smaller than 7.2 kb. Thus, PFGE and plasmid profiles are signatures of diversification in DT104.

Microarray analysis of *Salmonella* serovar Typhimurium DT104 isolates D1 to D13. In order to obtain information at a genome-wide level, we performed a comparative genome analysis using microarrays and DNA prepared from the 13 DT104 isolates. In this microarray analysis we focused on 4,110 array features representing the chromosome of *Salmonella* serovar Typhimurium DT104 strain NCTC13348, including the pSLT plasmid in this strain. The overwhelming finding of this analysis was the high degree of similarity between the *Salmonella* serovar Typhimurium DT104 isolates that displayed distinct PFGE patterns. The hierarchical clustering of the data showed some specific differences at the chromosomal level (Fig. 3). Prophage 1, prophage 2, and prophage 4 were present in all 13 *Salmonella* serovar Typhimurium DT104 isolates. However, differences were detected in the two other prophages. All of prophage 3 (plus the 5′ flanking chromosomal region) was undetectable in isolates D10 and D11. Consequently, PCR primers were designed to generate a DNA fragment spanning the insertion site of prophage 3 in the DT104 strain NCTC13348 genome, and sequencing of the DNA fragments generated using DNA prepared from isolates D10 and D11 indicated that no prophage or other DNA was inserted in the *att* sites in these isolates (data not shown). Interestingly, both these isolates came from the Aberdeen region and were isolated in the same year, but they were obtained from sporadic cases in different individuals. Prophage 5 was not present in isolates D7 and D8 but was present in the other 11 isolates. Again, PCR primers were designed to generate a DNA fragment spanning the insertion site of prophage 5 in the DT104 strain NCTC13348 genome, but in this case it was not possible to generate the fragment, indicating that the site is most likely occupied by a large DNA insertion, possibly an unknown phage (results not shown). Isolates D7 and D8 originated from different parts of Scotland.

We previously designed simple multiplex PCR-based assays for detecting variation in *Salmonella* serovar Typhimurium res-
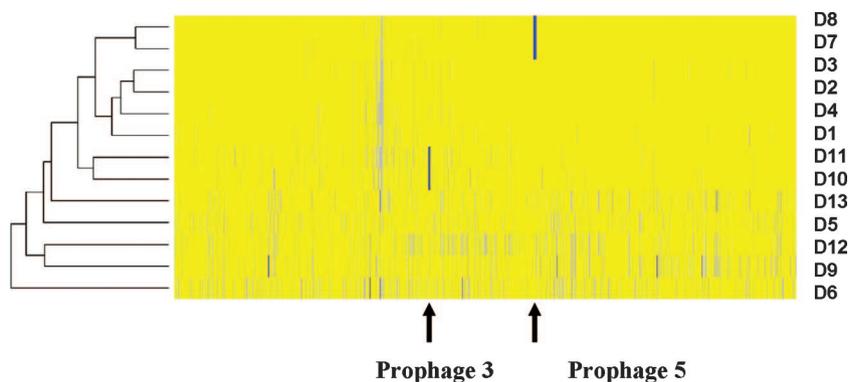
FIG. 3. Hierarchical clustering of microarray data for 13 *Salmonella* serovar Typhimurium DT104 isolates. The microarray data were processed by GACK for the SDT chromosomal loci of 13 isolates of *Salmonella* serovar Typhimurium DT104. Each row shows the results for a test isolate, as indicated on the right. The test/reference ratios were assessed to determine presence, absence, or uncertainty using GACK software, and the input data set was restricted to the 4,110 chromosomal features expected to be present in one or more of the isolates. The data are plotted in physical order of the SDT loci, and the results are indicated as follows: yellow, present; blue, absent or divergent; gray, unreliable as determined by the GACK software. The positions of prophages 3 and 5 are indicated.
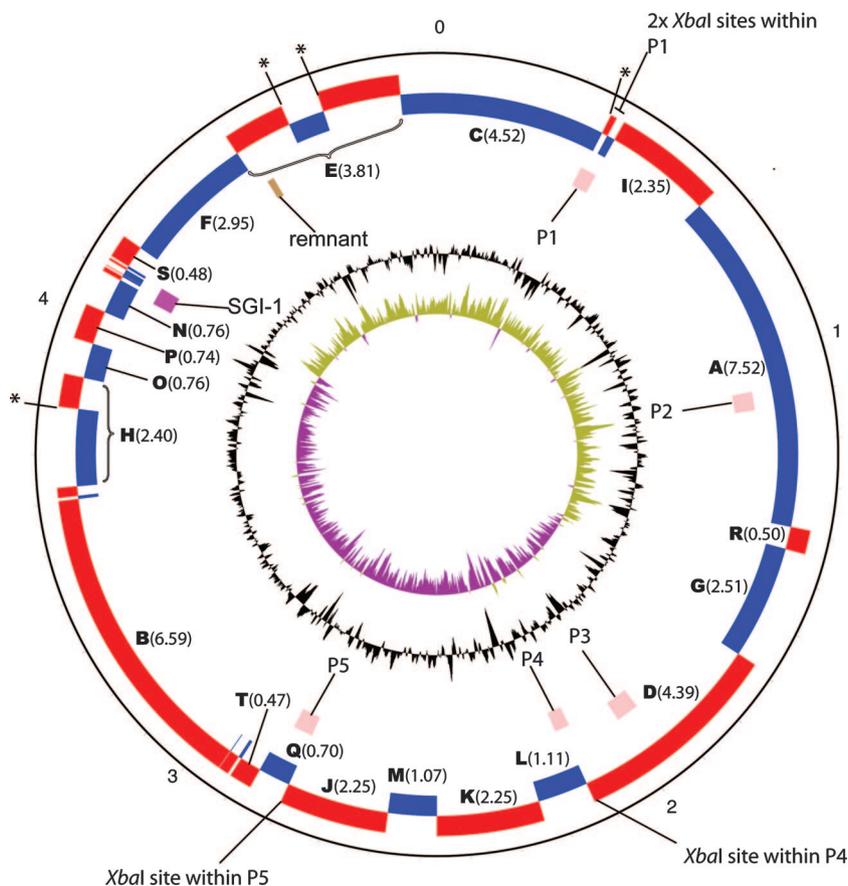
FIG. 4. Physical map of the *Salmonella* serovar Typhimurium NCTC113348 genome. The numbers outside the outer circle indicate the positions (in Mb) in the genome. Alternating inner blue and red bars indicate the putative XbaI fragments generated by restriction enzyme cleavage with this enzyme. Red and blue indicate adjacent fragments (drawn to scale). Individual fragments are labeled fragments A to T in order of size, starting with the largest fragment, fragment A; fragment sizes are indicated in parentheses (10^5 bp). Asterisks indicate Dam methylated XbaI sites, and brackets join fragments likely to run as a single band on a PFGE gel after XbaI cleavage of *Salmonella* serovar Typhimurium wild-type strain NCTC113348. The positions of prophages are indicated by pink rectangles, and the phage remnant is labeled. SGI1 is indicated by a purple rectangle. The inner black circle indicates the G+C content of the genome, whereas the inner green and purple circle indicates the GC skew.

ident prophages (7). Since the primers were designed using DT104 strain NCTC13348 as the template, multiplex PCRs were employed to detect prophages 1 to 5 and SGI1 in order to validate the microarray results for isolates D1 to D13. Only three different PCR profiles were generated, and these assays confirmed that prophage 3 was not present in D10 and D11 and prophage 5 was not present in D7 and D8 (data not shown). Aside from the prophages, other features that varied in the 13 *Salmonella* serovar Typhimurium isolates tended to be related to small DNA sequences rather than to groups of genes, and many of these features were restricted to individual isolates and therefore were not confirmed by sequencing.

Absence of prophage is reflected by the PFGE profile. We generated an in silico XbaI restriction map of the whole genome of *Salmonella* serovar Typhimurium DT104 strain NCTC113348 that could be used to predict the physical map locations of some of the DNA fragments visualized following PFGE of XbaI-cleaved genomic DNA (Fig. 4). This map could potentially be exploited to explain aspects of the PFGE patterns generated for the different DT104 isolates studied. Figure 4 shows the larger predicted XbaI DNA fragments that

might be visible on a gel (designated fragments A to T in order of diminishing size). Some XbaI target sites with distinct DNA sequence signatures are potentially subject to DAM methylation (Fig. 4), and consequently they may be masked on purified bacterial DNA and protected from XbaI (30). This would clearly account for differences between the observed and virtual PFGE patterns. Consequently, we generated a *dam*-negative mutant of NCTC13348 designated *Salmonella* serovar Typhimurium strain RAK102 (Δ*dam*::*aph*) and subjected DNA isolated from this derivative to XbaI cleavage (Fig. 5). Although the in silico predicted pattern had significant similarities to the experimental XbaI cleavage patterns generated using DNA from *Salmonella* serovar Typhimurium wild-type strain NCT13348 or RAK102 (Δ*dam*::*aph*), there were still differences (Fig. 5) (our unpublished observations). These differences may have been due to epigenetic factors involving DNA modification reactions. Nevertheless, we were in a position to partially interpret the PFGE patterns generated for the different DT104 isolates in the context of the physical map.

Initially, we focused on the prophage differences identified above. In silico analysis predicted that *Salmonella* serovar Ty-
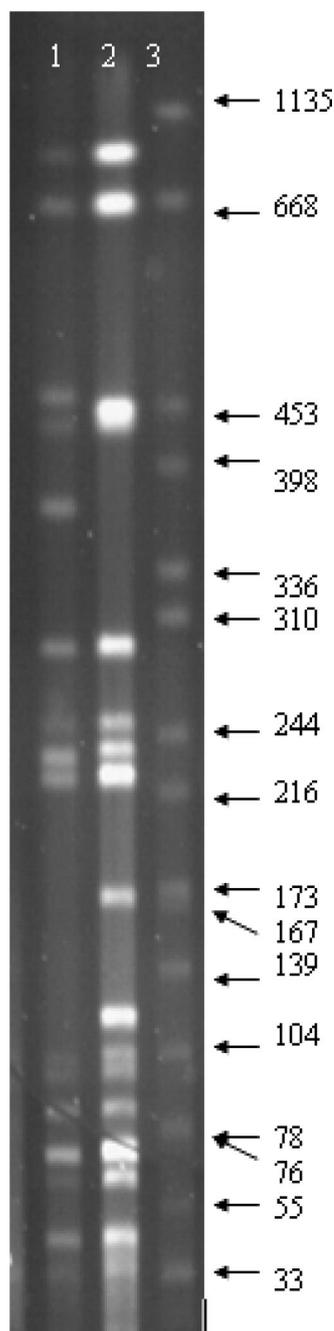
FIG. 5. PFGE profile generated after XbaI cleavage of DNA isolated from a *Salmonella* serovar Typhimurium DT104 strain NCTC13348 *dam*-negative mutant. Lane 1, DNA from wild-type strain NCTC13348; lane 2, DNA from *Salmonella* serovar Typhimurium DT104 *dam*-negative mutant (RAK102); lane 3, *Salmonella* serovar Braenderup marker. The band sizes indicated on the right are the band sizes for *Salmonella* serovar Braenderup.

phimurium DT104 strain NCTC13348 prophage 3 does not harbor an XbaI restriction site and is located on a 439-kb XbaI fragment located between genome coordinates 11691856 and 2131336 (Fig. 2 and 4, band D). DNA from all DT104 isolates which harbored prophage 3 generated a 439-kb band D that was a similar size on PFGE gels after XbaI cleavage. DNA

prepared from isolates D10 and D11, in which prophage 3 was not present, did not generate the 439-kb band D, but a novel 389-kb fragment was detected (Fig. 2, fragment D′), which was approximately 41 kb smaller, the size of prophage 3. Prophage 5 was predicted to have an XbaI restriction site near the 5′ end of the phage DNA at position 2798740. Thus, in silico analysis of *Salmonella* serovar Typhimurium DT104 strain NCTC13348 indicated that prophage 5 is located on two separate PFGE fragments that are 225 kb long (positions 2573535 to 2798740) and 70 kb long (positions 2798740 to 2868952) (fragments J and Q, respectively). Microarray data and multiplex PCR demonstrated that prophage 5 was not present in isolates D7 and D8. Therefore, these two isolates should not possess the restriction site located at position 2798740, and the 70- and 225-kb fragments which would be generated on each side of the cut site were predicted to be absent. Instead, XbaI cleavage of DNA from isolates D7 and D8 was expected to generate a 246-kb fragment, which was equivalent to coordinates 2573535 to 2868952 (295 kb) minus the size of the prophage 5 (49 kb). Examination of the sizes of the DNA fragments generated by XbaI cleavage of DNA prepared from isolates D7 and D8 indeed confirmed that 70-kb fragment Q was not present (Fig. 2). However, the predicted 246-kb fragment and 229-kb fragment J in these two isolates could not clearly be seen, as there were several DNA fragments in this region on the PFGE gel. To differentiate these bands, we synthesized DNA probes corresponding to genes located on the predicted 246-kb PFGE fragment and by Southern blotting confirmed the presence of this DNA fragment in the strains lacking prophage 5 (data not shown).

The impact of plasmids on the PFGE pattern was apparent since all isolates which harbored the virulence-associated plasmid pSLT generated a ~100-kb fragment on the PFGE gel. Previous work noted this link (29, 34). Isolates D5 and D6, which were pSLT negative, did not generate a fragment of this size (Fig. 2). Only one isolate (isolate D9) harbored a ~40-kb plasmid according to agarose gel analysis, and this isolate generated an additional ~33-kb fragment on a PFGE gel. Isolates D1, D2, and D11 produced an additional PFGE band at 53 to 56 kb and also harbored a ~60-kb plasmid. Plasmids less than 30 kb were more difficult to visualize by PFGE.

DISCUSSION

Previous studies using conventional typing approaches suggested that isolates of *Salmonella* serovar Typhimurium DT104 obtained from different sources were genetically highly related, although isolates can be distinguished by PFGE and variable number of tandem repeat analysis (7, 20). Using a whole-genome sequencing approach combined with DNA microarray analysis, we were able to more accurately define the genetic features of DT104 and highlight the specific variable regions of the genome undergoing relatively rapid evolution. The core genome of DT104 is highly conserved in different isolates. MLST revealed that all 13 isolates of *Salmonella* serovar Typhimurium DT104 used in this study were sequence type ST19. Inspection of the MLST *S. enterica* database revealed that 31 of 35 *Salmonella* serovar Typhimurium DT104 isolates at the time of writing were ST19 (http://web.mpiib-berlin.mpg.de /mlst/dbs/Senterica/). The four other sequence types reported,

ST40, ST153, ST159, and ST209, were each represented by one isolate. At the gene level the variation within our 13 DT104 isolates was limited mainly to two phage-like elements, SGI1, and various plasmids. Even when the sequenced DT104 strain NCTC13348 and strain LT2 isolated decades apart were compared, differences were restricted mainly to the same regions, aside from a previously defined deletion of genes involved in allantoin metabolism (22).

The 13 isolates of *Salmonella* serovar Typhimurium DT104 representing all PFGE patterns detected from human cases of gastroenteritis at the Scottish *Salmonella* Reference Laboratory between 2001 and 2005 were analyzed in detail. Together, these 13 isolates represented three different resistance types and had nine different plasmid profiles. Twelve isolates were multiply antibiotic resistant (ACSSuSpT), and four of these isolates also exhibited trimethoprim resistance. Subtypes of *Salmonella* serovar Typhimurium DT104 with R-type ACSSu SpTTm have been recognized in the United Kingdom since 1998 (35). These strains harbor a 6.8-kb plasmid (like the plasmid in isolate D2) with *dfrA14* encoding trimethoprim resistance, and plasmid acquisition may explain the different antibiotic resistance profiles expressed by different isolates. All isolates except isolates D5 and D6 harbored a ~100-kb plasmid consistent with the serovar-specific, virulence-associated plasmid designated pSLT in *Salmonella* serovar Typhimurium LT2 (SSP1 in DT104), and this was also confirmed by microarray analysis.

Microarray data for the 13 isolates indicated that they all harbored SGI1, prophage 1, prophage 2, and prophage 4. Interestingly, aside from plasmids, two main regions of variation were detected, both of which were situated within prophage-like elements. These regions were prophage 3, which was not present in isolates D10 and D11, and prophage 5, which was not present in isolates D7 and D8. All microarray findings were confirmed by multiplex PCRs. The *Salmonella* microarray detects prophages only from the sequenced genomes represented on the array, and consequently there could be other regions, likely to be prophage or other mobile elements, which are different in different isolates. Additional sequencing or other molecular approaches are required to identify such regions. Interestingly, the two prophages which varied, prophage 3 and prophage 5, had no obvious effect on the reaction to phage used in conventional phage typing, as all isolates were DT104. The finding that the prophage content varies in isolates of *Salmonella* serovar Typhimurium DT104 could be exploited in epidemiological typing. A multiplex PCR focusing on prophages 3 and 5 may prove to be useful for tracking isolates in outbreak situations. This PCR could quickly and easily determine whether isolates linked epidemiologically are similar or different at the genomic level.

By generating an in silico PFGE pattern for *Salmonella* serovar Typhimurium DT104 strain NCTC13348 and comparing this pattern with the actual XbaI patterns for the different isolates we were able to explain many of the band differences in terms of plasmids and prophages and made progress toward linking the PFGE pattern with the physical map of the genome. This approach could potentially be exploited in the future to regularly convert PFGE data into physical maps, facilitating a more detailed analysis of the phylogeny and genetic relatedness of different isolates. This study also describes differences in related *Salmonella* serovar Typhimurium isolates that appeared at different times and places during an epidemic, and our approaches could be utilized to define microevolution and the emergence of new variants or threats during an epidemic. We recognize that *Salmonella* serovar Typhimurium DT104 isolates not included in our studies, perhaps collected in other parts of the world, could exhibit other forms of variation, but our studies predict that such variation would most likely be associated with, or driven by, bacteriophages. Our analysis could serve as a basis for future comparisons.

Theoretically, similar approaches could be used for analysis of other *S. enterica* outbreaks or even outbreaks caused by other bacteria. Furthermore, our approach makes an important contribution to relating data from traditional typing protocols to whole-genome sequence data in order to more fully understand bacterial genome variation.

REFERENCES

1. **Abbott, J. C., D. M. Aanensen, K. Rutherford, S. Butcher, and B. G. Spratt.** 2005. WebACT—an online companion for the Artemis Comparison Tool. Bioinformatics **21:**3665–3666.
2. **Ahmed, A. M., A. I. Hussein, and T. Shimamoto.** 2007. *Proteus mirabilis* clinical isolate harbouring a new variant of Salmonella genomic island 1 containing the multiple antibiotic resistance region. J. Antimicrob. Chemother. **59:**184–190.
3. **Aleksic, S., F. Heinzerling, and J. Bockemühl.** 1996. Human infection caused by salmonellae of subspecies II to VI in Germany, 1977-1992. Zentbl. Bakteriol. **283:**391–398.
4. **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.
5. **Boyd, D., G. A. Peters, A. Cloeckaert, K. S. Boumedine, E. Chaslus-Dancla, H. Imberechts, and M. R. Mulvey.** 2001. Complete nucleotide sequence of a 43-kilobase genomic island associated with the multidrug resistance region of *Salmonella enterica* serovar Typhimurium DT104 and its identification in phage type DT120 and serovar Agona. J. Bacteriol. **183:**5725–5732.
6. **Carver, T. J., K. M. Rutherford, M. Berriman, M. A. Rajandream, B. G. Barrell, and J. Parkhill.** 2005. ACT: the Artemis Comparison Tool. Bioinformatics **21:**3422–3423.
7. **Cooke, F. J., J. Wain, M. Fookes, A. Ivens, N. Thomson, D. J. Brown, E. J. Threlfall, G. Gunn, G. Foster, and G. Dougan.** 2007. Prophage sequences defining hot spots of genome variation in *Salmonella enterica* serovar Typhimurium can be used to discriminate between field isolates. J. Clin. Microbiol. **45:**2590–2598.
8. **Datsenko, K. A., and B. L. Wanner.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. USA **97:**6640–6645.
9. **Doublet, B., D. Boyd, M. R. Mulvey, and A. Cloeckaert.** 2005. The Salmonella genomic island 1 is an integrative mobilizable element. Mol. Microbiol. **55:**1911–1924.
10. **Enright, M. C., and B. G. Spratt.** 1999. Multilocus sequence typing. Trends Microbiol. **7:**482–487.
11. **Farrant, J. L., A. Sansone, J. R. Canvin, M. J. Pallen, P. R. Langford, T. S. Wallis, G. Dougan, and J. S. Kroll.** 1997. Bacterial copper- and zinc-cofactored superoxide dismutase contributes to the pathogenesis of systemic salmonellosis. Mol. Microbiol. **25:**785–796.
12. **Figueroa-Bossi, N., and L. Bossi.** 1999. Inducible prophages contribute to Salmonella virulence in mice. Mol. Microbiol. **33:**167–176.
13. **Figueroa-Bossi, N., S. Uzzau, D. Maloriol, and L. Bossi.** 2001. Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in *Salmonella*. Mol. Microbiol. **39:**260–271.
14. **Frost, J.** 1994. Testing for resistance to antibacterial drugs, p. 73–82. *In* H. Chart (ed.), Methods in practical laboratory bacteriology. CRC Press, Boca-Raton, FL.
15. **Gaudriault, S., J. O. Thaler, E. Duchaud, F. Kunst, N. Boemare, and A. Givaudan.** 2004. Identification of a P2-related prophage remnant locus of *Photorhabdus luminescens* encoding an R-type phage tail-like particle. FEMS Microbiol. Lett. **233:**223–231.
16. **Gulig, P. A.** 1990. Virulence plasmids of *Salmonella typhimurium* and other salmonellae. Microb. Pathog. **8:**3–11.

17. **Kado, C. I., and S. T. Liu.** 1981. Rapid procedure for detection and isolation of large and small plasmids. J. Bacteriol. **145:**1365–1373.
18. **Kidgell, C., U. Reichard, J. Wain, B. Linz, M. Torpdahl, G. Dougan, and M. Achtman.** 2002. *Salmonella typhi*, the causative agent of typhoid fever, is approximately 50,000 years old. Infect. Genet. Evol. **2:**39–45.
19. **Kim, C. C., E. A. Joyce, K. Chan, and S. Falkow.** 2002. Improved analytical methods for microarray-based genome-composition analysis. Genome Biol. **3:**RESEARCH0065.
20. **Lawson, A. J., M. Desai, S. J. O'Brien, R. H. Davies, L. R. Ward, and E. J. Threlfall.** 2004. Molecular characterisation of an outbreak strain of multiresistant *Salmonella enterica* serovar Typhimurium DT104 in the UK. Clin. Microbiol. Infect. **10:**143–147.
21. **Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt.** 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. USA **95:**3140–3145.
22. **Matiasovicova, J., P. Adams, P. A. Barrow, H. Hradecka, M. Malcova, R. Karpiskova, E. Budinska, L. Pilousova, and I. Rychlik.** 2007. Identification of putative ancestors of the multidrug-resistant *Salmonella enterica* serovar Typhimurium DT104 clone harboring the Salmonella genomic island 1. Arch. Microbiol. **187:**415–424.
23. **McClelland, M., K. E. Sanderson, J. Spieth, S. W. Clifton, P. Latreille, L. Courtney, S. Porwollik, J. Ali, M. Dante, F. Du, S. Hou, D. Layman, S. Leonard, C. Nguyen, K. Scott, A. Holmes, N. Grewal, E. Mulvaney, E. Ryan, H. Sun, L. Florea, W. Miller, T. Stoneking, M. Nhan, R. Waterston, and R. K. Wilson.** 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. Nature **413:**852–856.
24. **Mmolawa, P. T., H. Schmieger, and M. W. Heuzenroeder.** 2003. Bacteriophage ST64B, a genetic mosaic of genes from diverse sources isolated from *Salmonella enterica* serovar Typhimurium DT 64. J. Bacteriol. **185:**6481–6485.
25. **Mulvey, M. R., D. A. Boyd, A. B. Olson, B. Doublet, and A. Cloeckaert.** 2006. The genetics of Salmonella genomic island 1. Microbes Infect. **8:**1915–1922.
26. **Peters, T. M., C. Maguire, E. J. Threlfall, I. S. Fisher, N. Gill, and A. J. Gatto.** 2003. The Salm-gene project—a European collaboration for DNA fingerprinting for. Eur. Surveill. **8:**46–50.
27. **Popoff, M. Y., and L. Le Minor.** 1992. Antigenic formulas of the *Salmonella* serovars, 5th ed. WHO Collaborating Center for Reference and Research on *Salmonella*, Institute Pasteur, Paris, France.
28. **Ridley, A., and E. J. Threlfall.** 1998. Molecular epidemiology of antibiotic resistance genes in multi-resistant epidemic *Salmonella typhimurium* DT 104. Microb. Drug Resist. **4:**113–118.
29. **Ridley, A. M., E. J. Threlfall, and B. Rowe.** 1998. Genotypic characterization of *Salmonella enteritidis* phage types by plasmid analysis, ribotyping, and pulsed-field gel electrophoresis. J. Clin. Microbiol. **36:**2314–2321.
30. **Sales, J., L. Vali, D. V. Hoyle, C. M. Yates, S. G. Amyes, and I. J. McKendrick.** 2007. The interaction between dam methylation sites and Xba1 restriction digest sites in *Escherichia coli* O157:H7 EDL933. J. Appl. Microbiol. **102:**820–825.
31. **Schmieger, H., and P. Schicklmaier.** 1999. Transduction of multiple drug resistance of *Salmonella enterica* serovar Typhimurium DT104. FEMS Microbiol. Lett. **170:**251–256.
32. **Swaminathan, B., T. J. Barrett, S. B. Hunter, and R. V. Tauxe.** 2001. PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. Emerg. Infect. Dis. **7:**382–389.
33. **Threlfall, E. J.** 2000. Epidemic *Salmonella typhimurium* DT 104—a truly international multiresistant clone. J. Antimicrob. Chemother. **46:**7–10.
34. **Threlfall, E. J., M. D. Hampton, H. Chart, K. L. Hopkins, L. R. Ward, and T. G. M.** 2004. Emergence of new subclones of multiresistant *Salmonella typhimurium* DT104 possibly associated with poultry meat. Vet. Rec. **154:**89–90.
35. **Threlfall, E. J., K. Hopkins, and L. Ward.** 2005. Diversification in *Salmonella typhimurium* DT104. Emerg. Infect. Dis. **11:**980–981.
36. **Threlfall, E. J., B. Rowe, J. L. Ferguson, and L. R. Ward.** 1986. Characterization of plasmids conferring resistance to gentamicin and apramycin in strains of *Salmonella typhimurium* phage type 204c isolated in Britain. J. Hyg. **97:**419–426.