

Genomic Comparisons of *Salmonella enterica* Serovar Dublin, Agona, and Typhimurium Strains Recently Isolated from Milk Filters and Bovine Samples from Ireland, Using a *Salmonella* Microarray

F. J. Reen,¹ E. F. Boyd,¹ S. Porwollik,² B. P. Murphy,^{3,4} D. Gilroy,⁴ S. Fanning,⁵
and M. McClelland^{2*}

Department of Microbiology, University College Cork, National University of Ireland-Cork,¹ Veterinary Department,³ and Department of Biological Sciences, Cork Institute of Technology,⁴ Cork, and Centre for Food Safety, Faculties of Agriculture, Medicine & Veterinary Medicine, University College, Belfield, Dublin,⁵ Ireland, and Sidney Kimmel Cancer Center, San Diego, California²

Received 4 June 2004/Accepted 21 October 2004

***Salmonella*-induced enterocolitis is the leading food-borne illness with a lethal outcome and causes millions of cases of gastroenteritis each year. We examined genomic variation among 12 environmental, veterinary, and clinical *Salmonella enterica* serovar Dublin, Agona, and Typhimurium strains isolated in Ireland between 2000 and 2003, as well as two clinical isolates from Canada and four archival isolates, which belonged to serovars Dublin and Agona. Using DNA-DNA hybridization to a microarray consisting of most of the predicted protein-encoding sequences of the *S. enterica* serovar Typhimurium LT2 genome, we identified a number of genomic regions that were absent in one or more serovars. The 34 genomic regions encoded proteins involved in sugar metabolism, transport, fimbrial and phage biogenesis, and transcriptional regulation, as well as inner and outer membrane-associated proteins. Two of the four prophages identified in strain LT2, prophages Gifsy-1 and Gifsy-2, were present in all six serovar Typhimurium strains examined. Prophage Fels-1 was absent from all 18 isolates examined, and Fels-2 was completely absent from the serovar Typhimurium isolates and the *Salmonella* Reference Collection B serovar Dublin strain Du2. All five *Salmonella* pathogenicity islands were present in all isolates. Plasmid pSLT was absent from all serovar Agona isolates, and only homologues of the *spv* genes were present in eight of the nine serovar Dublin strains. Only limited intraserovar diversity was found among the nine serovar Dublin, three serovar Agona, and six serovar Typhimurium isolates examined even though these isolates had extensive geographic, temporal, and source differences.**

Salmonella enterica is an important facultative intracellular pathogen of reptiles, birds, and mammals and is a serious food-borne pathogen of humans. *S. enterica* is implicated in a wide variety of life-threatening infections ranging from typhoid to gastroenteritis and bacteremia (20). *Salmonella*-induced enterocolitis is the leading food-borne illness with a lethal outcome in the United States, and it causes millions of cases of gastroenteritis each year (20). *Salmonella* strains are conventionally identified and classified according to the Kauffmann-White serotyping scheme, which is based on antigenic variation in the outer membrane lipopolysaccharide (O) and phase 1 (H1) and phase 2 (H2) flagella (8, 22, 23). To date, more than 2,400 serovars of *S. enterica* have been identified, and most serovars are capable of infecting a wide variety of animal species and humans. In addition, there is temporal and geographic variation in which serovars are most commonly associated with human infection. *S. enterica* serovar Typhimurium is among the serovars most commonly associated with human salmonellosis in most European countries (5, 6) and in the United States (<http://www.cdc.gov/ncidod/dbmd/phlisdata/Salmonella.htm>).

Serovar Typhimurium is frequently present in the gastrointestinal tracts of cattle, pigs, poultry, and other animal species

and is transferred to humans via the food chain. In particular, serovar Typhimurium definitive phage type 104 (DT104), R-type ACSSuT, has emerged as a global health problem (13). Serovar Dublin is most commonly associated with the gastrointestinal tract of cattle, but it can also infect humans. Serovar Agona was first isolated from cattle in Ghana (15) and emerged as a public health concern in the late 1960s in the United States, Europe, and Israel. In the mid-1990s, serovar Agona isolates were implicated in several outbreaks associated with ready-to-eat snacks manufactured in Israel (21).

Serotyping is a convenient and epidemiologically useful method for categorizing isolates, but it does not provide a basis for estimating evolutionary genetic relatedness among strains. Comparative genomic analysis with microarray technology is a powerful tool for the analysis of bacterial diversity and evolution, and it has been used in a number of studies to examine variation and to determine evolutionary relationships among the salmonellae (2, 4, 24, 29, 30).

In the present study, we utilized the *S. enterica* serovar Typhimurium LT2 microarray to identify genomic regions of divergence in a range of serovar Dublin, Agona, and Typhimurium isolates recovered from bovine and milk filter samples in southern Ireland. In addition, we compared these recent Irish isolates to archival bovine and clinical isolates belonging to the same serovars to determine whether there were temporal or geographic differences. We identified genomic regions that were present in some serovars but not others; however, the

* Corresponding author. Mailing address: Sidney Kimmel Cancer Center, 10835 Altman Row, San Diego, CA 92121. Phone: (858) 450-5990. Fax: (858) 550-3998. E-mail: mmcclelland@skcc.org.

TABLE 1. *Salmonella* strains used in this study

Strain	Serovar	Phage type	Source	Year	Locality	Reference
CIT-D124	Dublin		Clinical	2001	Cork	This study
CCC8	Dublin		Milk filter	2002	Cork	This study
CCC88	Dublin		Milk filter	2002	Cork	This study
C6531	Dublin		Bovine	2003	Cork	This study
C6548	Dublin		Bovine	2003	Cork	This study
Du1 (SARB12)	Dublin		Bovine	1986	Idaho	3
Du2 (SARB14)	Dublin				Thailand	3
Du3 (SARB13)	Dublin		Bovine	1982	France	3
DuA1	Dublin		Clinical	2001	Canada	26
CCC28	Agona		Milk filter	2002	Cork	This study
Ag1 (SARB1)	Agona				Peru	3
AgA1	Agona		Clinical	2002	Canada	26
C5391	Typhimurium	DT104	Feline	2000	Cork	This study
C5689	Typhimurium	DT104	Bovine	2000	Cork	This study
C847	Typhimurium	DT104	Bovine	2001	Cork	This study
C4633	Typhimurium	DT104	Avian	2002	Cork	This study
CCC2a	Typhimurium	DT104	Milk filter	2002	Cork	This study
C129	Typhimurium	DT104	Porcine	2003	Cork	This study

intraserovar diversity was limited. For the serovar Dublin isolates, only strain Du2, an isolate from Thailand recovered in the 1980s, diverged significantly from the other strains. Serovars Agona and Typhimurium were homogeneous.

MATERIALS AND METHODS

Bacterial strains. In this study, we examined 18 *S. enterica* isolates representing serovars Dublin, Agona, and Typhimurium (Table 1). Twelve of these isolates were recently recovered from milk filters and livestock on farms in the Munster region of southern Ireland between 2000 and 2003. The Regional Veterinary Laboratory strains included in this study were obtained from the Regional Veterinary Laboratory in Cork, Ireland, and were postclinical presentation isolates obtained from animals in the Cork region. In addition, isolates DuA1 and AgA1 (serovars Dublin and Agona, respectively) were recent clinical isolates from Canada. The remaining three serovar Dublin isolates and one serovar Agona isolate were obtained from *Salmonella* Reference Collection B (SARB) and were isolated in the 1980s (3).

Procedures for the detection of *Salmonella* spp. were carried out as recommended in the Food and Drug Administration *Bacteriological Analytical Manual* (19), with modifications. Briefly, filters were collected in a sterile container containing buffered peptone water (Oxoid CM509; Oxoid Ltd., Basingstoke, Hampshire, England). The filters were returned to the laboratory for analysis within 4 h of collection. The samples were processed, positive isolates were placed in serogroups by using the Wellcolex color test (Remel Inc., Remel Europe Ltd., Dartford, Kent, England), and the results were confirmed by using polyvalent O agglutinating sera (groups A to G) for *Salmonella* spp. (Murex Biotech Ltd., Kent, England). Biochemical confirmation was performed by using a bioMérieux API 20E test strip (bioMérieux SA, Marcy l'Etoile, France).

DNA extraction and probe labeling. Genomic DNA from a bacterial strain was prepared from 5 ml of a fresh overnight culture grown in Luria-Bertani broth at 37°C by using a QIAamp DNA extraction kit (QIAGEN, Valencia, Calif.) according to the manufacturer's instructions. Aliquots (1.5 µg) of harvested nucleic acid were labeled by using 12 µg of random hexamers, 2 nmol of Cy3- or Cy5-dCTP (Amersham, Piscataway, N.J.), and 10 U of Klenow enzyme (New England Biolabs, Beverly, Mass.) as previously described (30). Probes were purified by using a QIAquick PCR purification kit (QIAGEN) as suggested by the manufacturer, eluted in 1 mM Tris-HCl (pH 8.0), dried in a vacuum, and resuspended in 20 µl of sterile water. Immediately before use, the labeled probes of serovar Typhimurium strain LT2 (control sample) and one of the query strains (experimental sample) were mixed with equal volumes of 2× hybridization buffer containing 50% (vol/vol) formamide, 10× SSC, and 0.2% (wt/vol) sodium dodecyl sulfate and then denatured by boiling for 5 min (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate).

Microarray hybridization, image acquisition, and data analysis. Details concerning the construction and characteristics of the *Salmonella* microarray used in

this study have been described elsewhere (28, 29). Briefly, 97.5% of all 4,498 genetic elements annotated in *S. enterica* serovar Typhimurium LT2 are represented in triplicate on the array, as are 104 of the 109 annotated genetic elements of the virulence plasmid pSLT.

The genomic DNA of each strain was labeled with Cy3-dCTP and mixed with an equal amount of Cy5-dCTP-labeled genomic DNA of *Salmonella* serovar Typhimurium strain LT2 and then applied to the array. Probes were hybridized to the nonredundant *Salmonella* microarray overnight at 42°C in a hybridization chamber (Corning Glass Works, Corning, N.Y.) submerged in water. Protocols suggested by the manufacturer for hybridization in formamide buffer (http://www.corning.com/Lifesciences/technical_information/techDocs/gaps_ii_manual_protocol_5_02_cls_gaps_005.pdf) were used for prehybridization and posthybridization washing. Before scanning, the slides were dried by spinning them at 500 rpm for 5 min at room temperature. Scans were performed with a Scan Array 5000 laser scanner (7) by using the ScanArray 2.1 software (Packard BioChip Technologies). Signal intensities were quantified by using the QuantArray 3.0 software (Packard BioChip Technologies). Spots were analyzed by adaptive quantitation, an algorithm in the QuantArray software that compares the spot and median background intensities for each spot based on user-defined mask parameters for spot size and background diameter. The signal thresholds for spot and background are dynamic for every spot, which facilitates computational compensation for irregular spot shapes (<http://las.perkinelmer.com/content/Manuals/Quantarraymanual.pdf>). The local background intensity was subsequently subtracted from the recorded spot intensities. Ratios of the contribution of each spot to the total signal in each channel were calculated (data normalization). Negative values (i.e., the local background intensities were higher than the spot signals) were considered no data. Since the array was spotted in triplicate, one hybridization resulted in three data points per gene. The median of the three ratios for each gene was recorded. The lowest 5% of gene-specific signals were determined for the control sample (serovar Typhimurium strain LT2) and were excluded from the graphical representation in the figures.

The presence and absence of genes were evaluated based on a comparison of the normalized signal ratios of the query strain and serovar Typhimurium strain LT2 for the respective gene spots on the array. The median of the ratios for the genes in the STM genome flanked by elements with ratios higher than 0.67 on both sides (presence baseline P) and the standard deviation (SD_P) of these ratios were calculated for each query strain. Similarly, the medians and standard deviations of genes in the STM genome flanked by elements with ratios lower than 0.5 were determined (absence baseline A and SD_A). Genes with ratios which were higher than the presence threshold set at 2 SD_P below the baseline P value were scored as present, whereas genes with ratios lower than the absence threshold set at 2 SD_A above the baseline A value were scored as absent. Genes that were outside these thresholds were scored as uncertain.

RESULTS

Genomic regions absent in serovar Dublin. We examined nine *S. enterica* serovar Dublin isolates. Four of these isolates (CCC8, CCC88, C6531, and C6548) were recovered from bovine and milk filter samples from farms in southern Ireland between 2001 and 2003 (Table 1). Strains CIT-D1124 and DuA1 were recent clinical isolates from Ireland and Canada, respectively. The three other serovar Dublin isolates examined (Du1, Du2, and Du3) were from SARB and were bovine and human isolates from the United States, Thailand, and France, respectively, that were recovered in the 1980s (3). Comparative genomic analysis of these nine serovar Dublin isolates with the sequenced serovar Typhimurium strain LT2 revealed 12 regions that were absent among these isolates. These 12 regions were designated regions I, II, IV, XII, XIII, XV, XXIII, XXVI, XXVII, XXIX, XXX, and XXXIV (Table 2 and Fig. 1). The variable regions were primarily composed of genes encoding proteins involved in sugar metabolism and transport and fimbrial and phage biogenesis, as well as a number of putative cytoplasm and membrane proteins. Three of the 12 regions (regions II, IV, and XXVI) were absent only from Du2; 2 of the 12 regions (regions XII and XIII) were absent only from CIT-D124; and one region (region XXVII) was absent only from strains C6531 and CCC88. STM0081 to STM0084 (region I) (Fig. 1), which encoded two putative secreted proteins and a putative sulfatase, was absent in all serovar Dublin isolates except SARB isolates Du2 and Du3. The three regions shown to be absent only in Du2 (regions II, IV, and XXVI) contained the *stf* fimbrial operon, an enzymatic gene cluster, and a membrane-associated gene cluster, respectively. The region spanning genes STM1545 to STM1554 (part of region XIII) (Fig. 1), which encoded a number of cytoplasmic and regulatory proteins, was absent in all Dublin strains examined except SARB isolate Du2. In addition, serovar Dublin strain CIT-D124 had apparently undergone a deletion upstream of region XIII, STM1531 to STM1533 (region XII), which contained a hydrogenase gene cluster. Region XXIX (STM4195 to STM4218), which contained a number of phage genes, was absent from all nine serovar Dublin isolates examined (Fig. 1).

The *stc* fimbrial operon from STM2149 to STM2152 (region V) (Fig. 1) was absent in all serovar Dublin isolates examined. In addition, the *stjBC* genes of the *stj* fimbrial operon (region XXXIV) were also absent in all Dublin strains. A sugar transport cluster spanning genes STM3779 to STM3784 (region XXIII) (Fig. 1) was absent from all serovar Dublin strains except SARB strain Du2. The rhamnose catabolic operon *rhaBADs* (STM 4044 to STM4048), region XXVII, was absent from two of the nine serovar Dublin strains examined, C6531 (a bovine isolate) and CCC88 (a milk filter isolate from Ireland). In addition, region XXX (STM4417 to STM4436), which encoded sugar metabolism and transport proteins, thiamine synthesis proteins, and a number of dehydrogenases, as well as the *Salmonella* pathogenicity island 2 (SPI-2) effector protein SrfJ, was absent from all serovar Dublin isolates examined.

Four of the 12 regions uniquely variable in the serovar Dublin isolates, regions XV, XXIX, XXX, and XXXIV, were absent from all serovar Dublin isolates examined. One area (part of region XXIII) was absent from all serovar Dublin isolates

except Du2, and region I was absent from all serovar Dublin isolates except strains Du2 and Du3. Strain Du2 appears to be an unusual serovar Dublin isolate since it differed from the other eight, relatively homogeneous strains.

Genomic regions absent in serovar Agona. Three *S. enterica* serovar Agona isolates were examined; strain CCC28 was recovered from a milk filter in Ireland, strain AgA1 was a clinical isolate from Canada, and strain Ag1 was a SARB clinical strain from Peru (Table 1). Eight regions (regions VIII, XI, XIV, XVI, XX, XXIV, XXVIII, and XXXIII) (Fig. 2) that were absent in these isolates were identified (Table 2). The genes identified as absent specifically from the serovar Agona strains encoded flavoproteins (region VIII), cytoplasmic and membrane proteins (regions XI and XIV), phage proteins (region XIV), members of the LysR family of transcriptional regulators (regions VIII, XX, and XXIV), galactonate metabolic proteins (region XXIV), and a phosphotransferase transport system (region XXVIII).

The deletion spanning genes STM1862 to STM1871 (region XIV) (Fig. 2), which included *pagOK* and PhoP-PhoQ-activated genes, as well as several phage genes and an integrase gene, was absent from all serovar Agona isolates examined. Gunn and colleagues previously suggested that sequences surrounding *pagK* may have been or are part of a mobile element (16). Region XVI (STM2230 to STM2243) was absent from all serovar Agona isolates. This region encoded several phage proteins and included the *oafA* gene, which is responsible for the O-antigen acetylation step that generates the serovar-specific antigen. Part of this region, STM2230 to STM2234, including *oafA*, was also absent in the Dublin strains examined in this study (Table 2). A gene cluster (STM3117 to STM3123) encoding arylsulfatase and its corresponding regulator, two LysR family transcriptional regulators, lyase, hydrolase, and oxidase was also absent in all serovar Agona strains examined (region XX). The galactonate operon *dgoAKRT* (STM3827 to STM3830) (part of region XXIV) was absent in all of the serovar Agona strains examined. Phosphotransferase sugar transport (PTS) system genes (STM1127 to STM1133), the pyruvate formate lyase genes *pflCD*, and the fructose transport genes *frwBCD* (STM4110 to STM4116) (regions XI and XXVIII) were also absent in all serovar Agona isolates, as was region XXXIII (STM4534 to STM4540), which encoded a restriction-modification system.

Genomic regions absent in both serovar Dublin and serovar Agona. There were 11 genomic regions (regions III, VI, VII, X, XVII, XIX, XXI, XXII, XXV, XXXI, and XXXII) that were absent only in serovar Dublin and Agona isolates (Table 2). All of the region from STM0266 to STM0294 (region III) was absent in all three serovar Agona strains and one SARB serovar Dublin strain (strain Du2). Several genes encoding cytoplasmic proteins (STM0277 to STM0279 and STM0292 to STM0294) in this region were absent in the serovar Dublin strains, although the rest of the region appeared to be intact. A PTS gene cluster, STM0571 to STM0577 (region VI), had been deleted in all the serovar Agona and Dublin strains (Fig. 1 and 2 and Table 2). STM0716 to STM0727 (region VII) (Table 2), which encoded five glycosyltransferases and a phage integrase, was absent in all serovar Agona and Dublin isolates except SARB strain Du2. Similarly, region XXI (STM3251 to STM3256), a gene cluster involved in sugar metabolism, was

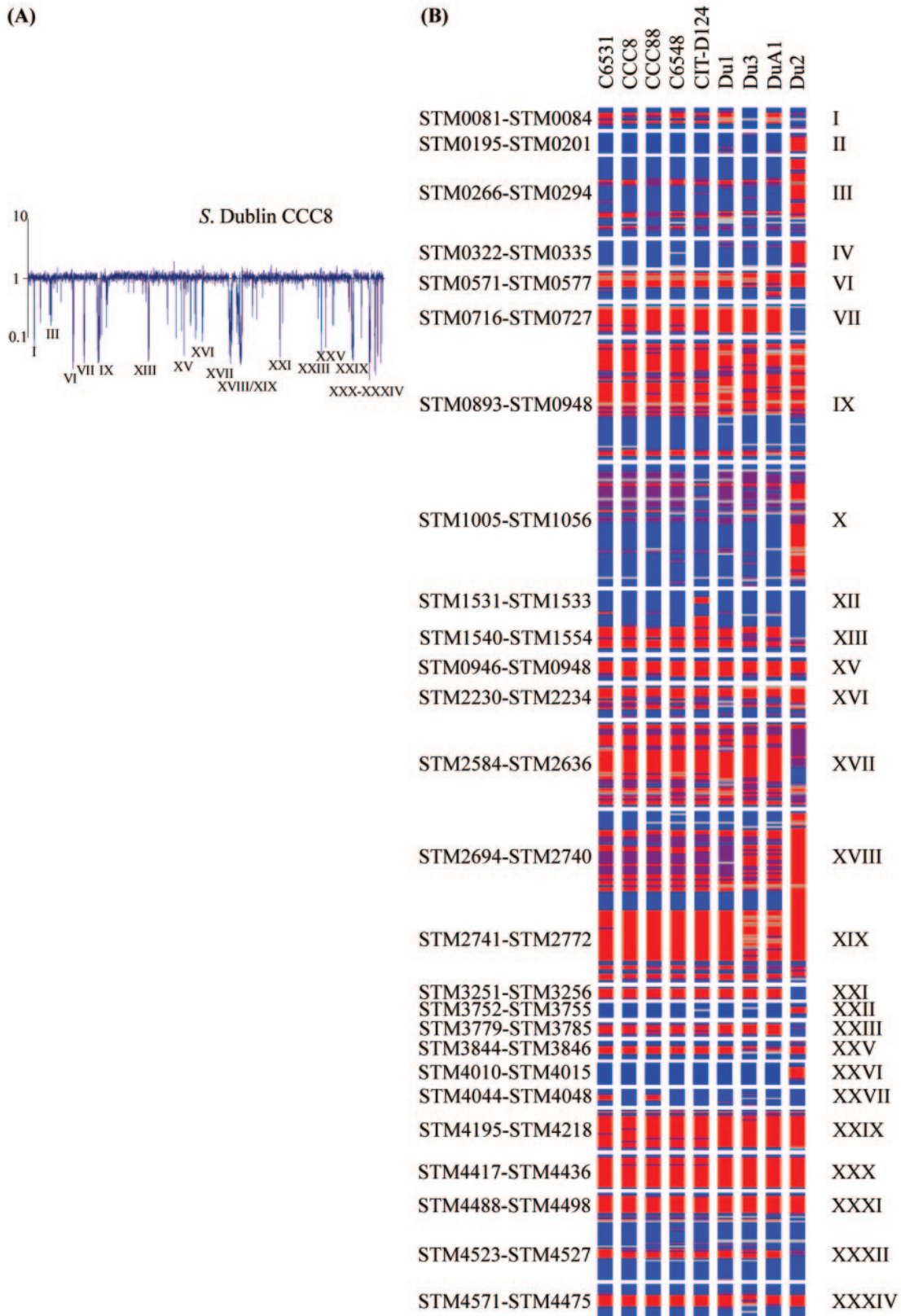


FIG. 1. Determination of regions absent in *S. enterica* serovar Dublin chromosomes. (A) Comparative genomic hybridization of an *S. enterica* serovar Dublin strain and serovar Typhimurium strain LT2. The representative plot shows the median signal ratios of the query *S. enterica* serovar Dublin strain over serovar Typhimurium strain LT2. Genes are plotted in order of their positions on the serovar Typhimurium strain LT2 genome. (B) Presence of *S. enterica* serovar Typhimurium LT2 protein-encoding sequence homologues in serovar Dublin strains. The gene status is color coded, as follows: blue, present; purple, uncertain; red, absent; and gray, low signal. Only the regions that harbor genes that are missing in serovar Dublin strains, along with some adjacent genes, are shown. The STM numbers on the left define these regions.

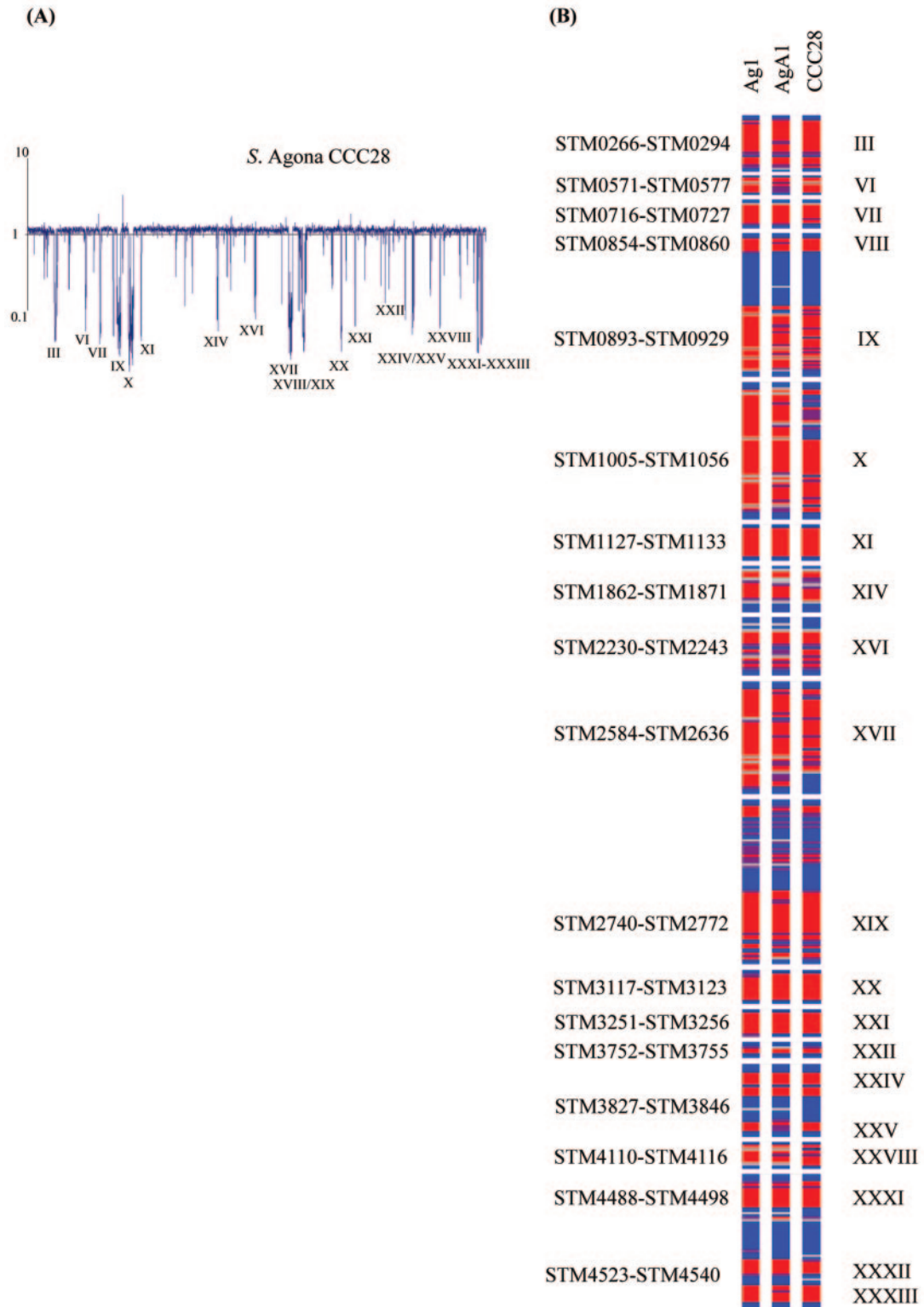


FIG. 2. Determination of regions absent in *S. enterica* serovar Agona chromosomes. (A) Comparative genomic hybridization of an *S. enterica* serovar Agona strain and serovar Typhimurium strain LT2. The representative plot shows the median of signal ratios of the query *S. enterica* serovar Agona strain over serovar Typhimurium strain LT2. Genes are plotted in order of their positions on the serovar Typhimurium strain LT2 genome. (B) Presence of *S. enterica* serovar Typhimurium LT2 protein-encoding sequence homologues in serovar Agona strains. The gene status is color coded, as follows: blue, present; purple, uncertain; red, absent; and gray, low signal. Only the regions that harbor genes that are missing in serovar Agona strains, along with some adjacent genes, are shown, and these regions are defined by their STM numbers.

absent in all serovar Agona and Dublin isolates except strain Du2 (Fig. 1 and 2).

Immediately downstream of the Fels-2 prophage, region XIX (STM2740 to STM2772), which encoded another PTS transport system and contained a sugar metabolism gene cluster, was absent from all the serovar Agona and Dublin isolates (Fig. 1 and 2 and Table 2). The 3' flanking end of this region contained three genes encoding transposases (STM2765, STM2768, and STM2769) in serovar Typhimurium strain LT2, and cross-hybridization likely resulted in the apparent presence of these genes in the serovar Agona and Dublin strains investigated. In addition, immediately downstream of the genes encoding the transposases the *fljAB* genes (STM2770 to STM2771) were found to be absent or divergent in all serovar Dublin and Agona strains, along with the highly conserved regulatory *hin* gene (STM2772) involved in flagellar phase switching. The absence of the *hin* gene results in monophasic isolates which have lost or never acquired the ability to switch phases.

STM2230 to STM2234, a subregion of region XVI, was absent in all the serovar Dublin and Agona isolates examined. STM4488 to STM4498 (region XXXI), a region flanked by a 5' integrase gene and composed primarily of cytoplasmic protein-encoding genes, as well as a restriction endonuclease gene and a methylase gene, was absent from all serovar Agona and Dublin strains examined (Fig. 1 and 2 and Table 2). A reverse transcriptase gene and an integrase gene (STM3844 to STM3846) (region XXV) downstream of the galactonate operon were absent from all serovar Agona strains, as well as all serovar Dublin strains. Region XXXII (STM4523 to STM4527), which contained several genes involved in sugar metabolism and transport, was absent from all serovar Agona and Dublin isolates examined (Table 2), except Du2.

Genomic regions absent in serovar Typhimurium. In this study, we examined six *S. enterica* serovar Typhimurium isolates recovered from four animal species (bovine, porcine, feline, and avian isolates) and from a milk filter in Ireland between 2000 and 2003. Only one region, region V (STM0517 to STM0529), was absent exclusively from the serovar Typhimurium isolates examined (Fig. 3). This region contained genes required for anaerobic nitrogen assimilation from allantoin and previously has been shown to be absent in the atypical monophasic *Salmonella* serovar [4,5,12:i:-], with the deletion affecting two-thirds of the glyoxylate carboligase (*gcl*) gene (12). This gene cluster was absent in all serovar Typhimurium strains examined in this study. However, the regulator of the allantoinase pathway, *allR*, was present in all strains.

The region encompassing genes STM3993 to STM4137 was amplified in serovar Typhimurium strain C5391 (Fig. 3). The amplification spanned the area between two rRNA clusters and was probably generated during homologous recombination between these clusters. The occurrence of an identical amplification has been reported for an archival serovar Typhimurium strain (28).

Prophages, *Salmonella* pathogenicity islands, and plasmid pSLT. In the present study, we found that Fels-1 (STM0893 to STM0929) (region IX) was missing in all 18 strains examined (Table 2). The Fels-2 prophage (STM2694 to STM2740) (region XVIII) was absent in all serovar Typhimurium strains and also in SARB strain Du2 (Table 2). However, Fels-2 prophage

was not completely deleted from the serovar Agona and Dublin strains, and both the 5' and 3' regions remained intact (Fig. 1 and 2). The serovar Agona strains retained a substantial portion of Fels-2, whereas the serovar Dublin strains appeared to have lost the central regions (STM2702 to STM2730). Gifsy-2 (STM1005 to STM1056) (region X) was present in all serovar Typhimurium strains and all serovar Dublin strains except Du2 (Table 2). The Gifsy-2 prophage was not present in the serovar Agona strains, with the exception of the region from STM1007 to STM1024, which was present in the milk filter isolate CCC28. Gifsy-2 has recently been reported to be at least partially present in several non-serovar Typhimurium *Salmonella* isolates (26, 29). The Gifsy-1 prophage (STM2584 to STM2636) (region XVII) was present in all six serovar Typhimurium strains but was absent from all serovar Agona and Dublin isolates (Table 2). However, a region of the Gifsy-1 prophage (STM2604 to STM2613) was present in serovar Dublin strain Du2.

To date, five SPIs have been identified (1, 25). We found that four of these five islands (SPI-1, SPI-2, SPI-4, and SPI-5) were present in all 18 isolates examined. However, the *sugR* and *rhuM* genes in SPI-3 (region XXII) were absent in all serovar Agona strains examined, as well as SARB strain Du2 (Table 2), but they were present elsewhere. Hansen-Wester and Hensel identified four additional DNA fragments (STM0557 to STM0586, STM2391 to STM2425, STM2227 to STM2247, and STM4301 to STM4324) containing genes known to be virulence factors that displayed characteristics of SPIs (17). All four regions were found to be present in each of the serovars examined in this study.

The *Salmonella* microarray also contained 104 of the annotated 109 genes of the strain LT2 virulence plasmid pSLT. In this study, we found that all of the serovar Agona strains, as well as the atypical serovar Dublin strain, strain Du2, lacked all pSLT genes, including the *spv* gene cluster. A 21-gene region containing the *spv* gene cluster was present in eight of the nine serovar Dublin strains. The entire plasmid was apparently present in all six serovar Typhimurium strains examined (data not shown).

Variable regions associated with tRNA genes, transposases, and integrases. Eight of the regions found to be variable in the 18 *S. enterica* isolates examined were associated with tRNA genes (Table 2). Two serovar Dublin-specific regions, region XV (four genes) and region XXX (20 genes), were associated with tRNA-MetG and L-seryl-tRNA^{Ser}, respectively. Two additional serovar Agona-specific variable regions, region XI (seven genes) and region XX (seven genes), were directly adjacent to the tRNA genes *serX* and *pheV*, respectively. Regions III (28 genes), XVI (14 genes), XXII (3 genes), and XXXI (10 genes) were directly adjacent to *aspV*, *pro2*, *selC*, and *leuX*, respectively (Table 2). Seven transposases and seven integrases were associated with deletions or insertions. For example, region III, deleted in the serovar Agona and Dublin isolates, was flanked on the 3' end by an integrase gene (Table 2). Six of the variable regions described in this paper (regions III, VII, XI, XVI, XIX, and XXXI) (Table 2) were designated a genomic island based on conformance to three of the following four criteria: aberrant G+C content, larger than 10 kb, associated with a tRNA gene, and flanked by an integrase or transposase (7).

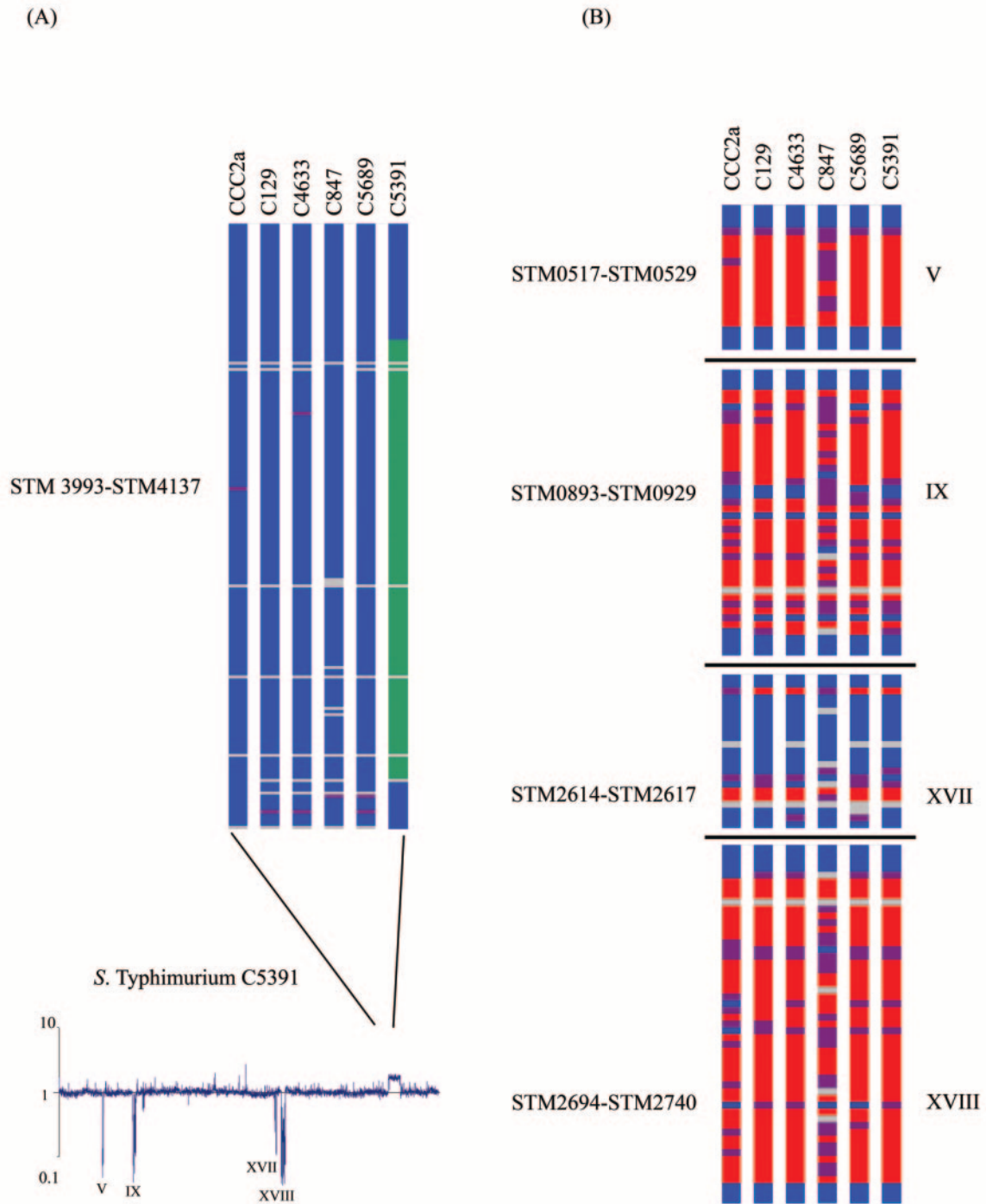


FIG. 3. Determination of regions absent in *S. enterica* serovar Typhimurium chromosomes. (A) Comparative genomic hybridization of an *S. enterica* serovar Typhimurium strain and serovar Typhimurium strain LT2. The representative plot shows the median of signal ratios of the query *S. enterica* serovar Typhimurium strain over serovar Typhimurium strain LT2. Genes are plotted in order of their positions on the serovar Typhimurium strain LT2 genome. The four regions containing genes apparently deleted from serovar Typhimurium strains are indicated, as is the amplified region in strain C5391. This amplification was not observed in the other five isolates examined. The apparent absence of region XVII is due to missing cross-hybridization with the absent Fels genes, resulting in lower ratios for the genes in this region. Therefore, these genes do not appear as absent in Table 2. (B) Presence of *S. enterica* serovar Typhimurium LT2 protein-encoding sequence homologues in other serovar Typhimurium strains. Gene status is color coded, as follows: blue, present; purple, uncertain; red, absent; green, amplified; and gray, low signal. Only the regions that harbor genes that are missing or amplified in serovar Typhimurium strains, along with some adjacent genes, are shown, and these regions are defined by their STM numbers.

DISCUSSION

We carried out a genome comparison of three serovars by using both clinical and environmental isolates from a variety of hosts, predominately from Ireland (Table 1). Considerable differences in the presence and absence of genes were observed between serovars, but little difference in gene content was observed within each serovar (Table 2). Limited genetic variation was observed among the six serovar Typhimurium strains; only strain C5391, a feline isolate, showed differences from all the other serovar Typhimurium isolates by harboring an *inter-rnm* amplification (Fig. 3). Similarly, among the serovar Agona isolates, the differences were limited to the partial presence of Gifsy-2 in strain CCC28 (Fig. 2). In the nine serovar Dublin isolates examined, apart from the atypical Du2 SARB isolate, which contained multiple variable regions when it was compared to the rest of the serovar Dublin isolates, there were only two variable intraserovar regions. The CIT-D124 clinical strain isolated in 2001 had lost a gene cluster containing PhoPQ-regulated genes, as well as genes encoding several hydrogenases (region XII and the 5' end of region XIII), and strains CCC88 and C6531 had lost the *rhaBADS* operon responsible for rhamnose metabolism (region XXVII). Recently, the serovar assignment of Du2 as a serovar Dublin isolate, although previously confirmed twice (3, 26), was questioned by a third serotyping effort (P. Fields, personal communication). These results suggest that the *S. enterica* serovar Typhimurium, Dublin, and Agona genomes have remained stable over the last two decades and that each serovar has wide geographic and host dissemination. Recently, Amavisit and coworkers suggested that although various *Salmonella* serovars cause a range of syndromes in both humans and animals, the natural distribution of insertion or deletion of some genes within the pathogenicity islands which they investigated was not significantly correlated with prominent differences in clinical features, host range, or levels of virulence (1). Perhaps the relevant differences are outside these islands.

A significant number of the regions identified in this study as being absent were involved in sugar metabolism and transport, highlighting the redundant nature of many of these gene clusters. Similarly, a number of fimbrial operons were absent; for example, the *stc* and *stj* fimbrial operons were missing in serovar Dublin isolates. To date, at least 18 different fimbrial operons have been identified in *S. enterica*, and the scattered phylogenetic distribution of fimbrial operons suggests that deletions and/or horizontal transfer of fimbrial operons occurred multiple times within *S. enterica* subspecies I, generating different combinations of fimbrial operons in different serotypes (27, 34).

The differences between *Salmonella* serovars are based on the surface antigen differences of O antigens, derived from the polysaccharide domain of the lipopolysaccharide, and H antigens, derived from the flagellin protein in the flagella. Most of the *fli* operon (STM1954 to STM1981) was predominantly present in all serovar Agona and Dublin isolates examined; the exception was *fliCD*, which were not detected in any of the serovar Dublin and Agona isolates, probably reflecting sequence divergence in these genes. For the phase I flagellin gene (*fliC*), comparative sequence analysis of multiple strains of *Salmonella* has strongly implicated horizontal gene transfer

and substitutive recombination as a major process generating new serovars (31–33). Much of the *Salmonella* O-antigen variation is a consequence of the extensive genetic diversity within *rfb* (O-antigen) gene clusters, which encode many of the enzymes involved in O-antigen biosynthesis and assembly (11). The O-antigen transferase *rfbX* gene is absent from, or divergent in, the serovar Dublin strains, along with a deletion and apparent partial deletion in the flanking genes, *rfbVJ*. These genes are present in all the serovar Agona strains. In serovar Typhimurium, the O antigen is acetylated, conferring the O5 serotype (18). The gene responsible for conferring acetylation on the O antigen, *oafA*, was absent in the serovar Agona and Dublin strains. The *oafA* gene is part of region XVI that encodes a number of phage proteins, as well as two homologues of the MsgA virulence protein. Downstream of *fliCD*, the *umuCD* error-prone repair genes also appeared to be absent in all serovar Agona and Dublin strains except strain Du2. These genes have been shown to play a critical role in *Escherichia coli* damage-inducible mutagenesis (14).

The regions of the genome that correspond to prophages have previously been reported to be variable across the serovars (4, 9, 10, 24, 26, 29). We found that the Fels-1 prophage was absent in all six serovar Typhimurium strains and also absent in all serovar Dublin and Agona isolates. The Fels-2 prophage was also absent in all serovar Typhimurium and Agona strains and was partially deleted in the serovar Dublin strains. The Gifsy prophages (except two Gifsy-1 genes) were predominantly present in all serovar Typhimurium strains, which is consistent with previous reports (9, 10, 29). Gifsy-1 was completely absent in all serovar Dublin and Agona strains, and Gifsy-2 was absent in serovar Agona strains but present in all serovar Dublin strains except strain Du2. *S. enterica* serovar Dublin is an example of a host-adapted serovar and is commonly associated with the bovine population. It is interesting that the Gifsy-2 prophage was present in the serovar Dublin isolates in this study. Both serovar Dublin and serovar Agona can survive in the gastrointestinal tract of bovines; therefore, one can expect a certain level of lateral gene transfer to occur. Several authors have highlighted the important role of bacteriophages in transmission of resistance markers originally located on the *Salmonella* chromosome. The observations reported here may highlight another example of this dynamic genetic flux.

The pSLT serovar Typhimurium virulence plasmid, which is required for systemic disease in the mouse (4), has previously been reported to be predominantly missing from all non-serovar Typhimurium serovars except serovars Dublin and Paratyphi (4). The three serovar Agona strains and serovar Dublin strain Du2 all lacked the 104 pSLT plasmid genes present on the array, including the *spv* gene cluster. Only the *spv* gene cluster was present in most of the other eight serovar Dublin strains examined; the exception was strain Du3, which harbored most of the plasmid.

Of the 34 regions which we identified as absent in some strains of our collection of isolates, 8 were associated with tRNAs, 14 were associated with integrases and/or transposases, and 15 had aberrant G+C contents (Table 2). Six of the 34 variable regions fulfilled at least three of four criteria for a genomic island, as defined by Dobrindt et al. (7). For example, region XXXI was a 21.4-kb region with a G+C content of

49.2% that was adjacent to a tRNA, and it encoded an integrase as well as restriction endonucleases. Region XIX was a 38.2-kb region inserted next to a tRNA with a G+C content of 48%, encoded an integrase, and contained a PTS gene cluster. Our data also show that *S. enterica* isolates contain a number of genomic islets (regions less than 10 kb long), which had several features of pathogenicity and metabolic islands.

These microarray studies identified unique features associated with serovar Dublin, Agona, and Typhimurium isolates. The data provide a comparison of environmental and clinical isolates belonging to these serovars with laboratory-adapted strains. The genomic features perhaps express characteristics necessary for host adaptation and virulence. The differences revealed may form the basis for a more detailed study of the role and importance of these markers, leading to a better understanding of the epidemiology and pathogenic potential in each case.

ACKNOWLEDGMENTS

This work was supported in part by NIH grants R01AI022933, R01AI52237, and R01AI34829 (to M.M.), by the generosity of Sidney Kimmel, and by an Enterprise Ireland International Collaboration grant (to E.F.B.).

We thank Eugene Power, Regional Veterinary Laboratory, Cork, Ireland, for providing several strains.

REFERENCES

- Amavisit, P., D. Lightfoot, G. F. Browning, and P. F. Markham. 2003. Variation between pathogenic serovars within *Salmonella* pathogenicity islands. *J. Bacteriol.* **185**:3624–3635.
- Boyd, E. F., S. Porwollik, F. Blackmer, and M. McClelland. 2003. Differences in gene content among *Salmonella enterica* serovar Typhi isolates. *J. Clin. Microbiol.* **41**:3823–3828.
- Boyd, E. F., F. S. Wang, 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.
- Chan, K., S. Baker, C. C. Kim, C. S. Detweiler, G. Dougan, and S. Falkow. 2003. Genomic comparison of *Salmonella enterica* serovars and *Salmonella bongori* by use of an *S. enterica* serovar Typhimurium DNA microarray. *J. Bacteriol.* **185**:553–563.
- Cormican, M., N. DeLappe, C. O'Hare, G. Doran, D. Morris, G. Corbett-Feeney, S. Fanning, M. Daly, M. Fitzgerald, and J. Moore. 2002. *Salmonella enterica* serotype Bredeney: antimicrobial susceptibility and molecular diversity of isolates from Ireland and Northern Ireland. *Appl. Environ. Microbiol.* **68**:181–186.
- Daly, M., J. Buckley, E. Power, C. O'Hare, M. Cormican, B. Cryan, P. G. Wall, and S. Fanning. 2000. Molecular characterization of Irish *Salmonella enterica* serotype Typhimurium: detection of class I integrons and assessment of genetic relationships by DNA amplification fingerprinting. *Appl. Environ. Microbiol.* **66**:614–619.
- Dobrindt, U., B. Hochhut, U. Hentschel, and J. Hacker. 2004. Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.* **2**:414–424.
- Ewing, W. H. 1986. Edwards and Ewing's identification of *Enterobacteriaceae*, 4th ed. Elsevier Science Publishing Co., Inc., New York, N.Y.
- Figuerola-Bossi, N., E. Coissac, P. Netter, and L. Bossi. 1997. Unsuspected prophage-like elements in *Salmonella typhimurium*. *Mol. Microbiol.* **25**:161–173.
- Figuerola-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.
- Fitzgerald, C., R. Sherwood, L. L. Gheesling, F. W. Brenner, and P. I. Fields. 2003. Molecular analysis of the *rfb* O antigen gene cluster of *Salmonella enterica* serogroup O:6,14 and development of a serogroup-specific PCR assay. *Appl. Environ. Microbiol.* **69**:6099–6105.
- Garaizar, J., S. Porwollik, A. Echeita, A. Rementeria, S. Herrera, R. M. Wong, J. Frye, M. A. Usera, and M. McClelland. 2002. DNA microarray-based typing of an atypical monophasic *Salmonella enterica* serovar. *J. Clin. Microbiol.* **40**:2074–2078.
- Gebreyes, W. A., S. Thakur, P. R. Davies, J. A. Funk, and C. Altier. 2004. Trends in antimicrobial resistance, phage types and integrons among *Salmonella* serotypes from pigs, 1997–2000. *J. Antimicrob. Chemother.* **53**:997–1003.
- Gonzalez, M., E. G. Frank, A. S. Levine, and R. Woodgate. 1998. Lon-mediated proteolysis of the *Escherichia coli* UmuD mutagenesis protein: in vitro degradation and identification of residues required for proteolysis. *Genes Dev.* **12**:3889–3899.
- Guinee, P. A., E. H. Kampelmacher, and H. M. Willems. 1961. Six new *Salmonella* types, isolated in Ghana (S. Volta, S. Agona, S. Wa, S. Techimani, S. Mampong and S. Tafo). *Antonie Leeuwenhoek* **27**:469–472.
- Gunn, J. S., W. J. Belden, and S. I. Miller. 1998. Identification of PhoP-PhoQ activated genes within a duplicated region of the *Salmonella typhimurium* chromosome. *Microb. Pathog.* **25**:77–90.
- Hansen-Wester, I., and M. Hensel. 2002. Genome-based identification of chromosomal regions specific for *Salmonella* spp. *Infect. Immun.* **70**:2351–2360.
- Hellerqvist, G. C., B. Lindberg, and S. Svensson. 1968. Structural studies on the O-specific side-chains of the cell-wall lipopolysaccharide. *Carbohydr. Res.* **8**:43–55.
- Hitchins, A. D., P. Feng, W. D. Watkins, S. R. Rippey, and L. A. Amaguana. 1998. *Bacteriological analytical manual*, 8th ed. Food and Drug Administration, Washington, D.C.
- Kennedy, M., R. Villar, D. J. Vugia, T. Rabatsky-Ehr, M. M. Farley, M. Pass, K. Smith, P. R. Cieslak, B. Imhoff, and P. M. Griffin. 2004. Hospitalizations and deaths due to *Salmonella* infections, FoodNet, 1996–1999. *Clin. Infect. Dis.* **38**(Suppl. 3):S142–S148.
- Killalea, D., L. R. Ward, D. Roberts, J. de Louvois, F. Sufi, J. M. Stuart, P. G. Wall, M. Susman, M. Schwieger, P. J. Sanderson, I. S. Fisher, P. S. Mead, O. N. Gill, C. L. Bartlett, and B. Rowe. 1996. International epidemiological and microbiological study of outbreak of *Salmonella* Agona infection from a ready to eat savoury snack. I. England and Wales and the United States. *BMJ* **313**:1105–1107.
- Le Minor, L. 1984. Genus III. *Salmonella* Lignieres 1900, 389^{AL}, p. 427–458. In N. R. Krieg and J. G. Holt (ed.), *Bergey's manual of systematic bacteriology*, vol. 1. Williams and Wilkins, Baltimore, Md.
- Le Minor, L., and J. Bockemuhl. 1988. 1987 supplement (no. 31) to the schema of Kauffmann-White. *Ann. Inst. Pasteur Microbiol.* **139**:331–335.
- 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.
- Ochman, H., and E. A. Groisman. 1996. Distribution of pathogenicity islands in *Salmonella* spp. *Infect. Immun.* **64**:5410–5412.
- Porwollik, S., E. F. Boyd, C. Choy, P. Cheng, L. Florea, E. Proctor, and M. McClelland. 2004. Characterization of *Salmonella enterica* subspecies I genovars using microarrays. *J. Bacteriol.* **186**:5883–5898.
- Porwollik, S., and M. McClelland. 2003. Lateral gene transfer in *Salmonella*. *Microbes Infect.* **5**:977–989.
- Porwollik, S., R. M. Wong, R. A. Helm, K. K. Edwards, M. Calcutt, A. Eisenstark, and M. McClelland. 2004. DNA amplification and rearrangements in archival *Salmonella enterica* serovar Typhimurium LT2 cultures. *J. Bacteriol.* **186**:1678–1682.
- Porwollik, S., R. M. Wong, and M. McClelland. 2002. Evolutionary genomics of *Salmonella*: gene acquisitions revealed by microarray analysis. *Proc. Natl. Acad. Sci. USA* **99**:8956–8961.
- Porwollik, S., R. M. Wong, S. H. Sims, R. M. Schaaper, D. M. DeMarini, and M. McClelland. 2001. The DeltavvB mutations in the Ames strains of *Salmonella* span 15 to 119 genes. *Mutat. Res.* **483**:1–11.
- Smith, N. H., P. Beltran, and R. K. Selander. 1990. Recombination of *Salmonella* phase 1 flagellin genes generates new serovars. *J. Bacteriol.* **172**:2209–2216.
- Smith, N. H., and R. K. Selander. 1991. Molecular genetic basis for complex flagellar antigen expression in a triphasic serovar of *Salmonella*. *Proc. Natl. Acad. Sci. USA* **88**:956–960.
- Smith, N. H., and R. K. Selander. 1990. Sequence invariance of the antigen-coding central region of the phase 1 flagellar filament gene (*fliC*) among strains of *Salmonella typhimurium*. *J. Bacteriol.* **172**:603–609.
- Townsend, S. M., N. E. Kramer, R. Edwards, S. Baker, N. Hamlin, M. Simmonds, K. Stevens, S. Maloy, J. Parkhill, G. Dougan, and A. J. Baumler. 2001. *Salmonella enterica* serovar Typhi possesses a unique repertoire of fimbrial gene sequences. *Infect. Immun.* **69**:2894–2901.