# Limited boundaries for extensive horizontal gene transfer among *Salmonella* pathogens

**Eric W. Brown, Mark K. Mammel, J. Eugene LeClerc, and Thomas A. Cebula†**

Division of Molecular Biology, Office of Applied Research and Safety Assessment (HFS-025), Center for Food Safety and Applied Nutrition, Food and Drug Administration, Laurel, MD 20708

Edited by Richard D. Kolodner, University of California at San Diego, La Jolla, CA, and approved October 24, 2003 (received for review July 15, 2003)

**Recombination is thought to be rare within *Salmonella*, as evidenced by absence of gene transfer among SARC strains that represent the broad genetic diversity of the eight primary subspecies of this common facultative intracellular pathogen. We adopted a phylogenetic approach to assess recombination within the *mutS* gene of 70 SARB strains, a genetically homogeneous population of *Salmonella enterica* subspecies I strains, which have in common the ability to infect warm-blooded animals. We report here that SARB strains show evidence for widespread recombinational exchange in contrast to results obtained with strains exhibiting species-level genetic variation. Besides extensive allele shuffling, SARB strains showed notably larger recombinagenic patch sizes for *mutS* (at least ≈1.1 kb) than previously reported for *S. enterica* SARC strains. Explaining these experimental dichotomies provides important insight for understanding microbial evolution, because they suggest likely ecologic and genetic barriers that limit extensive gene transfer in the feral setting.**

It becomes evermore apparent that horizontal gene transfer (HGT) underlies the mosaic structure of bacterial chromosomes. As much as a quarter of the genome of pathogenic *Escherichia coli* O157:H7, for instance, resides in islands of genes from donor species with a different base composition (1). Boundaries for HGT must exist, however, for unlimited exchange would obscure species boundaries within the bacterial kingdom. A comparison of *Salmonella enterica,* where HGT of *mutS* alleles was not evident among its eight subspecies (2), and *E. coli*, where extensive evidence for exchange of *mutS* alleles was found (3, 4), is noteworthy, as *Salmonella* is thought to be limited in recombinational exchange (5). A truly clonal family of organisms is ruled out in light of major DNA rearrangements, acquisitions, and losses recently uncovered among distinct *Salmonella* lineages (6, 7). Here, we show marked promiscuity among a homogeneous population (subspecies I) of *Salmonella* strains. Group I strains of *Salmonella* share a common niche, one restricted to warm-blooded mammals (8, 9), and they are likely endowed with compatible restriction-modification (R-M) systems that permit incorporation of longer segments of DNA. It appears that wholesale transfer of genes is possible among these strains, whereas such exchange is much more limited beyond the subspecies level.

In this study, we analyzed DNA sequences from the *mutS* gene, a key component of methyl-directed mismatch repair (MMR). Besides its role in mismatch correction, MutS also acts as a barrier to HGT by blocking recombination of diverged DNA (10). The association made recently between HGT and MMR gene evolution led to the hypothesis that exchange of *mutS* alleles might quiet the mutator phenotype caused by mutations in *mutS* (refs. 3 and 4; reviewed in ref. 11). The restoration of a functional *mutS* gene by recombination ensures both the long-term survival of the organism (12) and the simultaneous rescue of adaptive changes spawned by the *mutS* phenotype (13, 14). This hypothesis is supported both by phylogenetic data, showing that *mutS* is "scrambled" by recombination in *E. coli* (3, 4), and by direct genomic comparisons, revealing novel sequence insertions and rearrangements in the *mutS-rpoS* intergenic region of pathogenic *E. coli* and *Salmonella* strains (4, 15, 16). Moreover, the location of the SPI-1 pathogenicity-associated island directly adjoining *mutS* (2, 17) is further evocative that this region is subject to recombinational encounters. Curious, then, was our finding that only an isolated *mutS* recombinagenic patch of DNA in a single strain was observed across a diverse collection of strains [*Salmonella* reference collection C (SARC)], representative of the eight subspecies of *S. enterica,* suggesting that reassortment of alleles is limited in this region of the *Salmonella* chromosome (2). Because the SARC set comprises strains exhibiting species-level genetic variation and occupying unique ecologic niches (18, 19), these particular strains may be an exception and not generally indicative of *mutS* recombination across all *Salmonella* strains. To test this idea, we examined *mutS* evolution among a more genetically homogeneous set of *Salmonella* strains, the *Salmonella* reference collection B (SARB) comprising solely group I *Salmonella* pathogens (9).

## Materials and Methods

**Bacterial Strains and Culture Conditions.** A total of 70 bacterial strains, all group (subspecies) I *Salmonella* pathogens, was included as sources of DNA. All the *Salmonella enterica* group I strains used in this study originated from the SARB (*Salmonella* reference B) collection, kindly provided by E. F. Boyd, National University of Ireland, Cork, Ireland. SARB is recognized as representing the extent of genetic variability of *S. enterica* subspecies I (9). All *Salmonella* strains were cultured on LB Agar (Difco).

**PCR Amplification and Sequencing.** Genomic DNA was isolated by using a commercially available extraction matrix (Bio-Rad) according to the manufacturer's instructions. Oligonucleotide primer sequences and PCR conditions for amplification of the *S. enterica mutS* gene were used as described (2). The *mutS* gene segment amplified and sequenced corresponded to base pair coordinates 1771–2868 of the *mutS*-coding region in *Salmonella enterica* serotype Typhimurium (*S. enterica* group I) (GenBank accession no. M18965) and included the conserved ATP-binding domain, which lies in the COOH-terminal half of the *mutS* protein. Three sets of primer pairs were used to generate an 831-bp segment from the *mdh* gene in *S. enterica*. Primer pairs mdh1F–mdh1R, mdh2F–mdh2R, and mdh3F–mdh3R amplified overlapping segments of DNA that together spanned 831 bp, nearly the entire *mdh* locus in *Salmonella*. Primers were added to a final amount of 50 pmol and included: mdh1F, 5′-TCGGTCAGGCGCTGGCATTA-3′; mdh1R, 5′-CAGCTTACCTTTCAGCTCTGC-3′; mdh2F, 5′-TGGTGCAGCAGA-

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: HGT, horizontal gene transfer; ILD, incongruence length difference; MLEE, multilocus enzyme electrophoresis; MMR, methyl-directed mismatch repair; R-M, restriction-modification; SARC, *Salmonella* reference collection C; SARB, *Salmonella* reference collection B.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AY268620–AY268759).

†To whom correspondence should be addressed. E-mail: tcebula@cfsan.fda.gov.

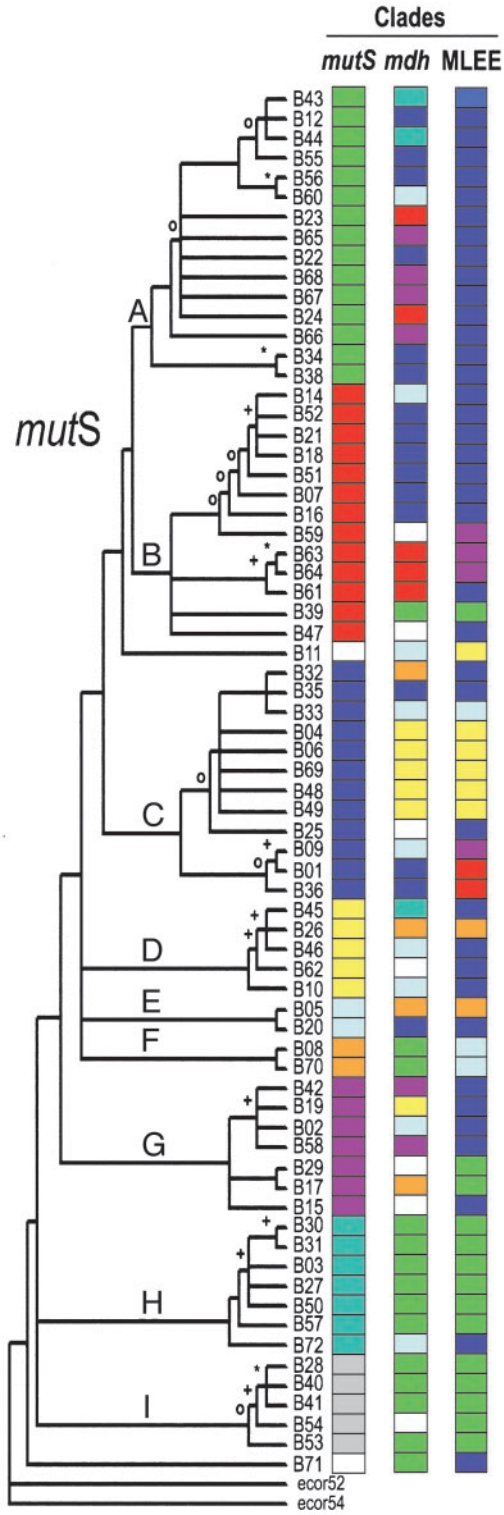


**Fig. 1.** Most parsimonious phylogenetic relationships of *S. enterica* group I *mutS* alleles. The tree shown resulted from a *mutS* multiple sequence alignment [CLUSTAL X (20)] that was analyzed phylogenetically by using PAUP* (21, 22). Nine *mutS* clades (designated as A–I) that contained multiple strains were identified from the tree. Distributions of these same strains within *mutS*, *mdh*, and MLEE clades are designated to the right of the *mutS* tree such that strains originating from the same *mutS*, *mdh*, and MLEE clades are depicted with a common color. Nodal support values in the form of bootstraps (5,000 iterations) are symbolized on the tree as follows: *, 76–100%; +, 51–75%; O, 26–50%; no symbol, 0–25%. The eight *mdh* clades presented here were obtained from maximum parsimony analysis as well. The seven MLEE clades were defined in a previous analysis of SARB strains (9). *mutS* and *mdh* trees were rooted by using *E. coli* as the outgroup (ECOR strains 52 and 64). Color cells that remain white represent those strains that could not be assigned to a specific multistrain clade for that data set. Disjunct distributions of color cells within the *mdh* and MLEE columns serve to illustrate the phylogenetic incongruence between these two markers of *Salmonella* chromosome evolution and *mutS*.

TCGCTAAAAC-3′; mdh2R, 5′-CCTTCCACATAGGCGCATTCC-3′; mdh3f, 5′-CAGAACGCCGGTACTGAAGTC-3′; and mdh3r, 5′-TCGGGCAGGAACAGCTTATTTAT-3′. PCR products were concentrated by using Qiaquick spin columns (Qiagen, Valencia, CA). Nucleotide cycle-sequencing was performed in both directions directly on purified PCR templates by automated Sanger dideoxy-chain termination methods and the primers described above (Amplicon Express, Pullman, WA).

**Phylogenetic Analysis.** Multiple-sequence alignment of the *mutS* nucleotide sequences was performed by using CLUSTAL X (20). Aligned nucleotide matrices were subjected to phylogenetic analysis by using the principle of maximum parsimony, available in PAUP* Version 4.03b (21). Most parsimonious trees were sought by using heuristic search methods with random addition of taxa and tree-bisection-reconnection in effect. A successive weighting strategy was applied to minimize the number of equally most-parsimonious cladograms (22), whereas combinable component (semistrict) consensus methods were used to coalesce most-parsimonious trees in such a way that only those strain relationships that were not in topological conflict among any of the original trees are represented (23).

Congruence between genes was assessed by using the incongruence length difference (ILD) test (24) (1,000 partitions) available in PAUP* Version 4.03b (21). Overall compatibility of sites was measured for combined and partitioned DNA sequences by using RETICULATE (25). Binary sites only (informative sites containing exactly two distinct nucleotides) were included in the compatibility of sites analysis. Genetic distances (Jukes–Cantor-corrected) between SARB and SARC *mutS* clades and overall diversity levels within the two populations were measured by using MEGA2 Version 2.1 (26).

## Results and Discussion

**The Evolutionary History of *Salmonella mutS* Is Distinct from That of the Whole Chromosome.** We adopted a phylogenetic approach to assess recombination within the *mutS* gene of 70 SARB strains. The most parsimonious phylogeny was constructed for these strains by using 1,098 bp of the 3′ half of the *mutS* gene (Fig. 1). The SARB *mutS* phylogeny was then compared with phylogenies derived from either multilocus enzyme electrophoresis (MLEE) (9) analysis or *mdh* gene sequences. Note that a phylogeny derived from DNA that has been acquired laterally would display incongruence (phylogenetic discordance) when compared with evolutionary trees constructed from stable housekeeping genes or whole-chromosome measures of diversity, such as MLEE (27). *mdh* encodes the glycolytic enzyme, malate dehydrogenase, and is one of several housekeeping genes whose evolutionary history appears to recapitulate the evolutionary histories of the *Salmonella* and *E. coli* chromosomes (28). As such, this gene has been used as an "anchor locus" in reiterating strain phylogeny for enteric species (2, 3, 28). Here, nine multistrain *mutS* clades (A–I) were identified from the tree; we assigned a unique color code so that each strain within the same clade shares a common color cell (Fig. 1). Likewise, eight *mdh* and seven MLEE clades (9), each containing multiple SARB strains, were identified from their respective phylogenies and also color-coded. It is important to note that the colored cells are unique to a phylogenetic data set; i.e., each column (data set) in Fig. 1 is uniquely color-coded.

EVOLUTION

On inspection, we found that strains composing single SARB lineages or clades based on *mdh* and MLEE phylogenies were distributed across disparate clades in the *mutS* phylogeny (Fig. 1). In every case, one or more strains from each of the eight *mdh* clades that contained multiple members were displaced into distinct *mutS* clades. For example, the single *mdh* clade containing SARB strains 23, 24, 63, 64, and 61, each coded red in Fig. 1 (*mdh* column), is dispersed between two different clades (A and B) on the *mutS* tree. This finding suggests that these particular strains, although linked tightly in *mdh* evolution, retain *mutS* alleles with distinct, and more unrelated, evolutionary histories. Similarly, strain(s) from six of the seven MLEE lineages were displaced into separate clades on the *mutS* tree, again indicating that *mutS* familial ties belie *Salmonella* strain relationships. The only exception was the MLEE lineage comprising SARB 01 and SARB 36 (depicted by red cells in the MLEE column of Fig. 1), where this grouping retained a single-clade structure in the *mutS* tree, except for the insertion of SARB 09.

**Congruence and Compatibility of the *mutS* vs. *mdh* Genes.** Discordance between *mutS* and *mdh* was tested further by using the ILD test, which evaluates the likelihood of a common evolutionary (congruent) history between different genes or distinct domains within the same gene (24). ILD methodology is rooted in cladistic analysis, thus providing consistency with the tree-building measures used here. In previous analyses of *S. enterica* gene sequences, we showed that the ILD analyses were supported by independent measures of compatibility analysis, split decomposition graphing, and maximum $\chi^2$ testing (2).

The two genes revealed significant incongruence by ILD ($P$ = 0.001 for 1,000 partitions) when all 70 SARB strains were analyzed. These data show a high degree of phylogenetic discordance between the 1.1-kb *mutS* segment analyzed here and the *S. enterica* chromosome, presumably reflecting the numerous HGTs of *mutS* alleles that have accumulated during the radiation of group I pathogens. Compatibility of sites (25) between *mutS* and *mdh* for all 70 SARB strains was consistent with ILD-based discordance for these two genes. Two sites are deemed compatible if they can be accounted for once in a phylogeny. Incompatible sites can be the result of either HGT or redundant mutations occurring at a single site (25). The *mdh* gene yielded an overall compatibility of 75.2% when evaluated with the 1.1-kb *mutS* sequence and a compatibility of 88.8% when analyzed separately. These findings are in stark contrast to our previous analysis of SARC strains, where recombination, observed in all but one *S. enterica* isolate, was limited to a 510-bp patch of sequence within the 1.1-kb *mutS* sequence analyzed above (2).

**Differences in the Extent and Length of *mutS* Exchange Among *Salmonella* SARC and SARB Strains.** Both tree-building and ILD approaches revealed recombination for the entire *mutS* segment in SARB. One possible explanation for these results is the disproportionate sampling of SARB ($n$ = 70) and SARC ($n$ = 16) strains. To negate this artifactual explanation, we analyzed our data, adopting a submatrix-sampling method whereby 100 sixteen-strain submatrices of the SARB strains were generated randomly and where each submatrix contained SARB strain samples from across the entire SARB *mutS* tree. Each *mutS* submatrix was then ILD-tested against a corresponding *mdh* submatrix consisting of identical strains. Of the 100 submatrices analyzed, only three yielded ILD scores >0.05 (1,000 partitions) (Fig. 2*A*). Thus, the vast majority (97%) of the 16-strain 1.1-kb submatrices was significantly incongruent with *mdh,* an indication that these data sets retained one or more SARB strains that were the recipients of horizontally transferred *mutS* sequences. Differences in *mutS* evolution between SARB and SARC are also supported by tree analysis (Fig. 2*B*). Although both populations revealed *mutS* patch segments (base pairs 424–933) that were phylogenetically distinct from *mdh* and from the *Salmonella* chromosome, perfect agreement between *mutS* and *mdh* was observed in the analysis of the entire 1.1-kb SARC sequence and for sequences flanking both sides of the patch (Fig. 2*B*). In contrast, the 1.1-kb *mutS* sequence from SARB 16-strain subset 95, for example, maintained a scrambled evolutionary pattern when compared with *mdh* whether *mutS* was inspected in its entirety or dissected into nonpatch (flanking) sequences (Fig.

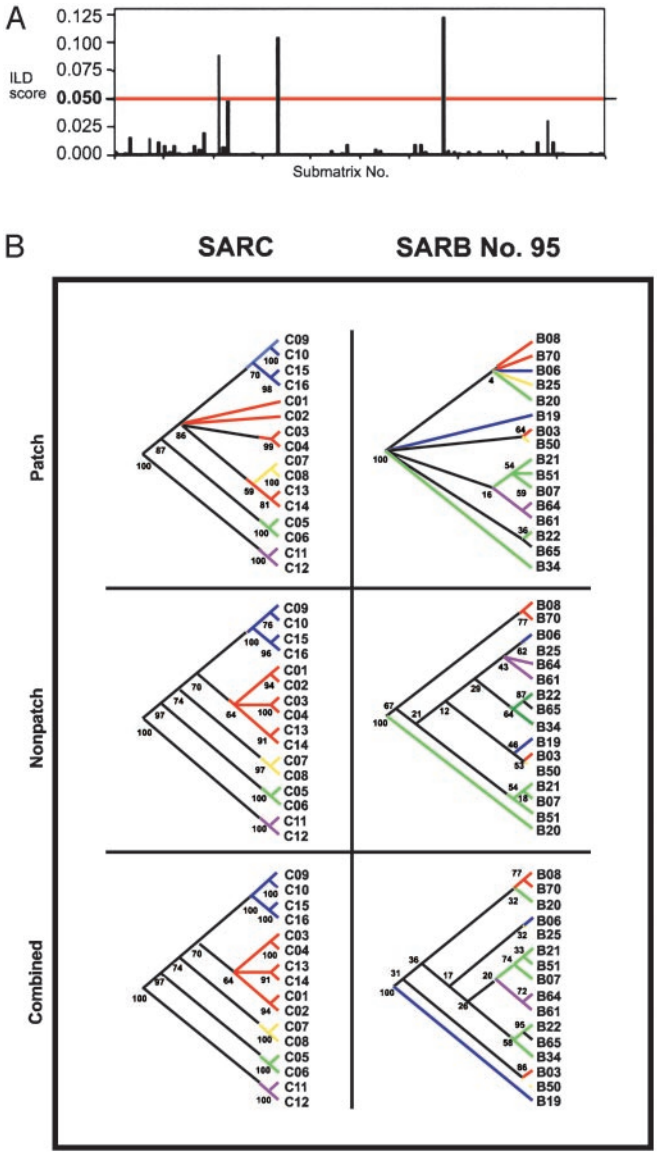


**Fig. 2.** Phylogenetic discordance between *S. enterica* group I *mutS* and *mdh* alleles. (*A*) Histogram displaying the ILD test (24) results for 100 sixteen-strain SARB submatrices that compared *mutS* and *mdh* for congruence to each other. The score for rejection of the null hypothesis of congruence ($P$ = 0.050) is denoted by a red line across the graph. (*B*) Comparison of 16-strain *mutS* phylogenies for the *S. enterica* SARC and SARB collections (submatrix 95). *mutS* trees shown resulted from the partitioned analysis of (*i*) the reported 510-bp horizontally transferred patch, (*ii*) the combined 5′ and 3′ flanking sequences surrounding the patch, and (*iii*) the total 1.1-kb *mutS* segment (2). Branches on the *mutS* tree are color-coded according to clades identified in the corresponding *mdh* trees. A black branch indicates a strain that did not cluster with any other strain in the *mdh* trees. Trees were rooted with two strains of *E. coli*. Bootstrap values are reported beside each respective node on the *mutS* trees.

**Table 1. Genetic diversity among *mutS* alleles in *S. enterica* SARC clades**

| | Mean nucleotide diversity ± SE, % | | | | | | |
|---|---|---|---|---|---|---|---|
| Clade (*n*) | II | IIIa | IIIb | IV | V | VI | VII |
| I (2) | 3.00 ± 0.48 | 7.50 ± 0.87 | 4.60 ± 0.66 | 4.50 ± 0.63 | 11.40 ± 1.25 | 2.20 ± 0.41 | 4.20 ± 0.60 |
| II (2) | 8.30 ± 0.95 | 5.50 ± 0.72 | 5.20 ± 0.66 | 12.40 ± 1.21 | 3.10 ± 0.47 | 5.00 ± 0.66 | |
| IIIa (2) | 7.10 ± 0.83 | 7.70 ± 0.89 | 11.40 ± 1.19 | 7.10 ± 0.84 | 7.30 ± 0.85 | | |
| IIIb (2) | 6.10 ± 0.79 | 12.30 ± 1.25 | 4.20 ± 0.58 | 6.00 ± 0.78 | | | |
| IV (2) | 12.90 ± 1.28 | 4.40 ± 0.62 | 1.90 ± 0.40 | | | | |
| V (2) | 11.30 ± 1.14 | 12.60 ± 1.26 | | | | | |
| VI (2) | 4.40 ± 0.62 | | | | | | |

Mean pairwise *mutS* sequence variation among the eight subspecies (clades) of the *S. enterica* SARC collection defined in Fig. 1. The distances shown are Jukes–Cantor-corrected and were calculated by using the program MEGA 2 (26).

2*B*). Moreover, SARB *mutS* segments (patch and nonpatch) from this 16-strain matrix yielded significant ILD values ($P < 0.05$) with *mdh* regardless of whether they were combined or partitioned [combined, $P = 0.001$; nonpatch, $P = 0.001$; patch, $P = 0.010$; 5′ of patch, $P = 0.002$; and 3′ of patch, $P = 0.024$]. Differences in tree structure and the ILD scores reflected by the submatrices analysis suggest that *mutS* recombination is pervasive among the SARB population and cannot be attributed to only a few recombinagenic transgressions.

**Clade Diversity Within *mutS* Is Inversely Correlated With the Extent of *mutS* Recombination.** Differences in the incidence of horizontally transferred *mutS* alleles between the two *Salmonella* collections may be reflected in the genetic diversity present within these sets. Whereas SARC contains several genetically disparate members (18, 19), SARB strains all emanate from *Salmonella* subspecies I (9). Accordingly, SARB strains should exhibit substantially less genetic variation at the nucleotide level. A survey of *mutS* clade diversity shows that the average genetic distance that exists between distinct clades differs significantly ($P < 0.001$) within the SARB and SARC populations (Tables 1 and 2). Mean interclade diversity between the eight SARC clades, subdividing the eight subspecies of *Salmonella* (9), was 6.9% with a maximum diversity approaching 13% [8.3% with the highly divergent *Salmonella bongori* strains (subspecies V) omitted]. In contrast, the nine SARB clades (Fig. 1) displayed a mean interclade diversity of <1% (0.94%) with a maximum diversity of 1.37% (between clades A and E). Likewise, overall diversity (II) of SARC strains was 6.49% (4.85% without *S. bongori*), whereas for SARB strains, II was 0.88%. The mean within-clade diversity was roughly equal for the two populations (SARC, 0.45% and SARB, 0.42%), demonstrating that our measure was not biased by dissimilar intraclade diversity measures for the two groups of strains. The observed disparity in clade diversity is particularly intriguing given observations made here that SARB strains appear to exchange *mutS* alleles far more readily than their SARC counterparts, a finding consistent with studies correlating recombination rate with the extent of genetic diversity (29).

**Mosaicism of the *mutS* Gene Within *Salmonella* Subspecies I.** Numerous exchanges of the 1.1-kb *mutS* segment within SARB strains, contrasted with all but the single example of *mutS* mosaicism within SARC strains (2), suggested that intragenic recombination partitioned this 1.1-kb sequence into subgenomic segments with distinct lineages. To investigate this possibility, we partitioned the total SARB matrix of 70 strains, and each of the 100 sixteen-strain submatrices as well, into three intragenic segments based on coordinates for the recombinagenic patch in *mutS* that was discerned by using the ILD test and a sliding-window approach (patch, base pairs 424–933; 3′ flank, base pairs 1–423; and 5′ flank, base pairs 934–1098) (2). Each segment was then ILD tested against the remaining two segments to reveal levels of congruence. Surprisingly, the total SARB matrix, and 42% of the submatrices, revealed at least one significant ILD partition [defined by either a red (0.010–0.001) or yellow (0.050–0.011) cell (Fig. 3*A*, column I)]. An additional 13% of the submatrices yielded at least one of three partitions that approached significance, displaying ILD scores from 0.100 to 0.051 as indicated by green cells in Fig. 3*A*. Significant ILD scores detected in this congruence array implicate intragenic HGT in the structuring of the 1.1-kb *mutS* sequence. In addition to the total *mutS* matrix, 2 submatrices (SARB 40 and SARB 73) were distinguished in having significant ILD values for all three of the segmental comparisons, and 10 other submatrices (13, 18, 23, 25, 49, 55, 60, 63, 79, and 96) maintained significant ILD scores for at least two of the three comparisons. Compatibility of sites analysis further supported these findings (Fig. 3*B*). The entire 1.1-kb *mutS* sequence for all 70 SARB strains yielded a compatibility of 74.1%. Two of the three intragene partitions, the patch sequence

**Table 2. Genetic diversity among *mutS* alleles in *S. enterica* SARB clades**

| | Mean nucleotide diversity ± SE, % | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Clade (*n*) | B | C | D | E | F | G | H | I |
| A (15) | 0.99 ± 0.21 | 0.96 ± 0.23 | 1.02 ± 0.24 | 1.37 ± 0.29 | 0.96 ± 0.27 | 1.00 ± 0.22 | 1.12 ± 0.26 | 1.24 ± 0.29 |
| B (13) | 0.92 ± 0.21 | 0.92 ± 0.21 | 1.20 ± 0.25 | 0.82 ± 0.23 | 0.93 ± 0.20 | 0.90 ± 0.66 | 1.00 ± 0.25 | |
| C (12) | 0.80 ± 0.19 | 0.88 ± 0.20 | 0.56 ± 0.20 | 0.72 ± 0.19 | 0.82 ± 0.22 | 1.06 ± 0.28 | | |
| D (5) | 1.18 ± 0.23 | 0.68 ± 0.21 | 0.86 ± 0.20 | 0.86 ± 0.21 | 1.08 ± 0.27 | | | |
| E (2) | 0.82 ± 0.21 | 1.04 ± 0.23 | 1.15 ± 0.25 | 1.36 ± 0.29 | | | | |
| F (7) | 0.60 ± 0.18 | 0.76 ± 0.23 | 0.84 ± 0.26 | | | | | |
| G (7) | 0.64 ± 0.16 | 1.01 ± 0.26 | | | | | | |
| H (5) | 0.85 ± 0.25 | | | | | | | |

Mean pairwise *mutS* sequence variation among the nine SARB clades defined in Fig. 1. The distances shown are Jukes–Cantor-corrected and were calculated by using the program MEGA2 (26).

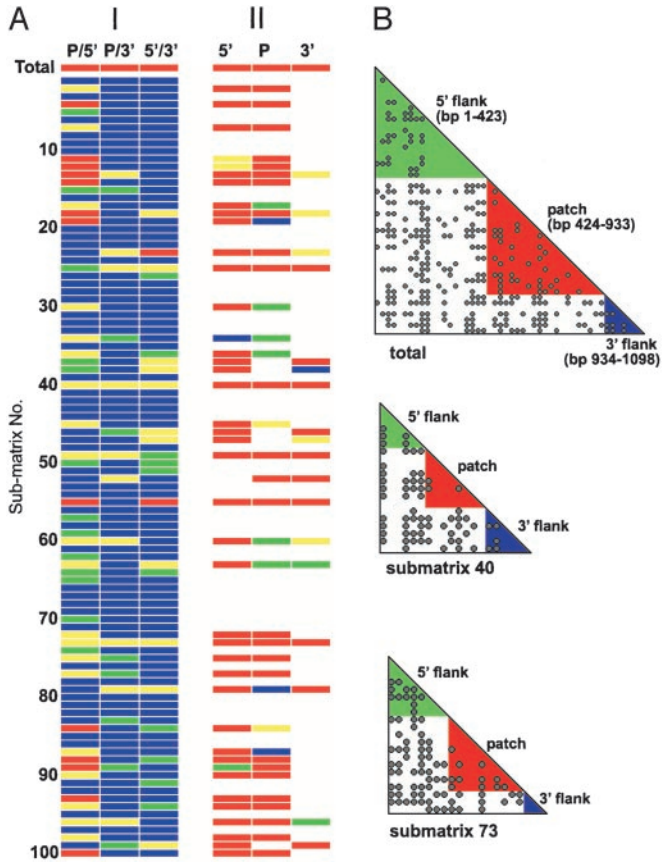


**Fig. 3.** Phylogenetic evidence for structural mosaicism of the group I *S. enterica mutS* gene. (*A*) Column I, congruence array of the three intragenic segments composing the 1.1-kb *mutS* sequence for the total *mutS* matrix and the 100 sixteen-strain submatrices. Intragene comparisons are noted at the top of each column (P, patch; 5′, 5′ flanking sequence; and 3′, 3′ flanking sequence). ILD scores are represented as individual color cells and are organized into four distinct ranges: red, 0.001–0.010; yellow, 0.011–0.050; green, 0.051–0.100; and blue, 0.101–1.00. Column II, ILD comparisons of the three intragenic segments to *mdh* for the total *mutS* matrix and 100 submatrices. (*B*) Compatibility matrix of the total *mutS* matrix showing pairwise comparisons of informative binary sites within (colored triangles) and between (white) the intragene segments indicated. Labels on the diagonal denote the 1.1-kb segment and corresponding base pair coordinates being compared. The matrices shown were constructed in a program written in C++ by M.K.M. for determination and visualization of incompatible sites. The algorithm is similar to that described in the program RETICULATE (25).

and 5′ flanking sequence, displayed improved compatibilities of 82.4% and 82.6% when analyzed alone. In general, compatibilities for submatrices 40 and 73 also improved when individual *mutS* segments were analyzed separately [submatrix 40, 65.3% (entire 1.1-kb sequence), 89.3% (patch), 66.7% (5′ flank), and 66.7% (3′ flank); submatrix 73, 59.5% (entire 1.1-kb sequence), 75.6% (patch), 57.1% (5′ flank), and 100% (3′ flank). Within-segment improvements are illustrated by a reduction in incompatible sites (depicted here as black dots) within the colored triangles of each matrix; red, green, and blue triangles denote compatibility comparisons inside of the patch, 5′ flank, and 3′ flank, respectively (Fig. 3*B*). Analyzed as a whole, these data point to substantial levels of incongruence within *mutS* of SARB strains, drawing attention to gene mosaicism within this population of *Salmonella*.

**Multiple Crossovers Have Forged the *mutS* Gene of *S. enterica* Subspecies I.** To investigate congruence between these discontinuous *mutS* segments and the *Salmonella* chromosome, segments responsible for significant ILD results among the total *mutS* matrix and the 42 significant *mutS* submatrices were individually tested against a matching matrix composed of corresponding *mdh* sequences (Fig. 3*A*, column II). In most tests, including the total *mutS* matrix, *mutS* segments that were incongruent with each other were also incongruent with *mdh*, suggesting that multiple crossovers have forged *mutS* structure in many of the strains within the collection of group I pathogens. Specifically, 71% ($n = 30$) of the 42 sixteen-strain submatrices retained multiply incongruent 1.1-kb sequences when the three intragenic partitions were tested against the *mdh* gene. In these matrices, both of the partitioned segments that contributed to a significant ILD score when compared with one another (Fig. 3*A*, column I) also yielded significant scores against *mdh* (Fig. 3*A*, column II). For example, 33 of the 42 submatrices with significant intragenic partitions retained significant tests between the patch (base pairs 424–933) and the 5′ flanking sequence (base pairs 1–423). Of these 33 tests, 24 showed both segments to be incongruent with *mdh*. This pattern held true as well for most of the patch to 3′ flank and 5′ flank to 3′flank tests (Fig. 3*A*). These results paint a more complex picture for HGT within the *mutS* gene of *Salmonella* group I pathogens than the one depicted for the diversity of *Salmonella* strains (SARC) as a whole (2). It is noteworthy that of the 56 intragenic ILD tests displaying significant incongruence within the 42 submatrices (Fig. 3*A*, column I), nearly 60% ($n = 33$) involved comparison of the patch to the flanking 5′ sequence, and of these, 73% were tests where both segments were significantly incongruent with *mdh* (Fig. 3*A*, column II). This finding implies that HGT of advantageous DNA into the *mutS* region of the *Salmonella* chromosome could, but need not, be limited to the *mutS* active site or even the *mutS* gene itself.

**Niche Overlap and R-M Compatibility as Potential Factors Delimiting Transfer of *mutS* Alleles in *Salmonella*.** Like previous reports of *mutS* genetic exchange among feral *E. coli* strains (3, 4), our results demonstrate that HGT has been pervasive during the evolution of *mutS* alleles among group I *Salmonella* pathogens. HGT of *mutS* sequences occurs much more frequently among genetically homogenous populations (e.g., a largely panmictic group of SARB strains) than among the more genetically diverged SARC strains (2, 9). It appears that a genetic threshold exists, one that tolerates free exchange of *mutS* sequences within a framework delimited by sequence variation and niche diversity of individual strains. HGT of *mutS* sequences between more distantly related strains would then be expected to be rare (2, 29, 30). Thus, the model for *mutS* recombination in *Salmonella* that emerges is one marked by largely unrestricted HGT of *mutS* alleles among closely related group I (SARB) pathogens but with only limited exchange beyond the subspecies level (e.g., exchange between SARC groups I, II, IIIa, IIIb, and IV–VII).

Enhanced HGT, which stems from MMR defects, may underscore the prominence of MMR mutators in the evolution of *Salmonella* pathogen populations (13). Recombination appears to have scrambled the *mutS* gene among many *Salmonella* group I pathogens, possibly one result of selection that would mitigate the deleterious effects of the hypermutable phenotype of *mutS* individuals in nature (2–4). The shuffling of *mutS* alleles would in fact be hastened by *mutS* defects, because wild-type MutS is known to inhibit homeologous recombination in *Salmonella* (30). Opportunities for this level of exchange likely materialized as a result of the common niche that *S. enterica* subspecies I share. These salmonellae are usually isolated from warm-blooded animals, whereas other non-group I strains are isolated generally from reptiles (8, 31). Our further observation that the entire 1.1 kb of *mutS* and flanking sequences appear to have

recombined among group I *Salmonella* also points to compatible R-M complexes that would permit the successful transfer of larger gene segments among closely related *Salmonella* pathogens; crosses between strains with identical R-M systems would not be subject to restriction (32). A gradation in the size limits of DNA segments exchanged likely exists that depends on the polymorphic character of R-M systems in natural strains (33). In sum, R-M compatibility and niche sharing may define genetic and ecologic boundaries that have seemingly limited recombination among the whole of *Salmonella*, but not within the group I subspecies. Evidence that levels of *mutS* recombination reflect these boundaries earmarks the *mutS* gene as a "molecular gauge" in assaying HGT in natural populations of bacterial pathogens.

We dedicate this work to the memory of Dr. Philip E. Hartman. We acknowledge A. Perlloni for technical assistance, M. Winkler, M. Kotewicz, D. Levy, and J. Lopez for insightful comments, and B. Barnett for assistance with figures. We thank E. F. Boyd for kindly contributing the SARB and SARC strain collections.

1. Perna, N. T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J., Kirkpatrick, H. A., *et al.* (2001) *Nature* **409,** 529–533.
2. Brown, E. W., Kotewicz, M. L. & Cebula, T. A. (2002) *Mol. Phylogenet. Evol.* **24,** 102–120.
3. Denamur, E., Lecointre, G., Darlu, P., Tenaillon, O., Acquaviva, C., Sayada, C., Sunjevaric, I., Rothstein, R., Elion, J., Taddei, F., *et al.* (2000) *Cell* **103,** 711–721.
4. Brown, E. W., LeClerc, J. E., Li, B., Payne, W. L. & Cebula, T. A. (2001) *J. Bacteriol.* **183,** 1631–1644.
5. Feil, E. J., Holmes, E. C., Bessen, D. E., Chan, M.-S., Day, N. P. J., Enright, M. C., Goldstein, R., Hood, D. W., Kalia, A., Moore, C. E., *et al*. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 182–187.
6. Liu, S. L. & Sanderson, K. E. (1998) *FEMS Microbiol. Lett.* **164,** 275–281.
7. Porwolik, S., Wong, R. M. & McClelland, M. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 8956–8961.
8. Baumler, A. J. (1997) *Trends Microbiol.* **5,** 318–322.
9. Boyd, E. F., Wang, F.-S., Beltran, P., Plock, S. A., Nelson, K. & Selander, R. K. (1993) *J. Gen. Microbiol.* **139,** 1125–1132.
10. Matic, I., Rayssiguier, C. & Radman, M. (1995) *Cell* **80,** 507–515.
11. Ochman, H. (2001) *Curr. Opin. Genet. Dev.* **11,** 616–619.
12. Funchain, P., Yeung, A., Stewart, J. L., Lin, R., Slupska, M. M. & Miller, J. H. (2000) *Genetics* **154,** 959–970.
13. LeClerc, J. E., Li, B., Payne, W. L. & Cebula, T. A. (1996) *Science* **274,** 1208–1211.
14. Cebula, T. A. & LeClerc, J. E. (1997) *Bull. Inst. Past.* **95,** 97–106.
15. Kotewicz, M. L., Brown, E. W., LeClerc, J. E. & Cebula, T. A. (2003) *Trends Microbiol.* **11,** 2–6.
16. LeClerc, J. E., Li, B., Payne, W. L. & Cebula, T. A. (1999) *J. Bacteriol.* **181,** 7614–7617.
17. Groisman, E. A. & Ochman, H. (1997) *Trends Microbiol.* **5,** 343–349.
18. Boyd, E. F., Wang, F.-S., Whittam, T. S. & Selander, R. K. (1996) *Appl. Environ. Microbiol.* **62,** 804–808.
19. Reeves, M. W., Evins, G. M., Heiba, A. A., Plikaytis, B. D. & Farmer, J. J., III (1989) *J. Clin. Microbiol.* **27,** 313–320.
20. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997) *Nucleic Acids Res.* **25,** 4876–4882.
21. Swofford, D. L. (1999) PAUP* (Smithsonian Inst., Washington, DC), Version 4.03b.
22. Carpenter, J. M. (1994) *Cladistics* **10,** 215–220.
23. Forey, P. L., Humphries, C. J., Kitching, I. L., Scotland, R. W. Siebert, D. J. & Williams, D. M. (1992) *Cladistics: A Practical Course in Systematics* (Clarendon Press, Oxford).
24. Farris, J. S., Kallersjo, M., Kluge, A. G. & Bult, C. (1995) *Cladistics* **10,** 315–319.
25. Jakobsen, I. B. & Eastal, S. A. (1996) *CABIOS* **12,** 291–295.
26. Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. (2001) MEGA2 (Pennsylvania State Univ., State College).
27. Dykhuizen, D. E. & Green, L. (1991) *J. Bacteriol.* **173,** 7257–7268.
28. Boyd, E. F., Nelson, K., Wang, F.-S., Whittam, T. S. & Selander, R. K. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 1280–1284.
29. Roberts, M. S. & Cohan, F. M. (1993) *Genetics* **134,** 401–408.
30. Zahrt, T. C. & Maloy, S. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 9786–9791.
31. Kraus, A., Guerra-Bautista, G. & Alarcon-Segovia, D. (1991) *J. Rheumatol.* **18,** 1328–1331.
32. Milkman, R. (1997) *Genetics* **146,** 745–750.
33. Bullas, L. R., Colson, C. & Neufeld, B. (1980) *J. Bacteriol.* **141,** 275–292.