APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Mar. 1996, p. 804–808 Vol. 62, No. 3
0099-2240/96/$04.00+0

Molecular Genetic Relationships of the Salmonellae

E. FIDELMA BOYD,† FU-SHENG WANG, THOMAS S. WHITTAM, AND ROBERT K. SELANDER*

Institute of Molecular Evolutionary Genetics, Mueller Laboratory, Pennsylvania State University, University Park, Pennsylvania 16802

Received 5 September 1995/Accepted 29 December 1995

A multilocus enzyme electrophoresis analysis of 96 strains of the salmonellae distinguished 80 electrophoretic types (ETs) and placed them in eight groups, seven of which correspond precisely to the seven taxonomic groups (I, II, IIIa, IIIb, IV, V, and VI) previously defined on the basis of biotype and genomic DNA hybridization. In addition, multilocus enzyme electrophoresis identified an eighth distinctive group (designated VII) composed of five strains that had been assigned to group IV on the basis of biotype. An analysis of variation in the combined nucleotide sequences of five housekeeping genes among 16 strains representing all eight groups yielded estimates of overall genetic relationships that are fully consistent with those indicated by DNA hybridization. However, the nucleotide sequences of seven invasion genes (*inv*/*spa*) in the strains of group VII were closely similar to those of strains of group IV. These findings are interpreted as evidence that group VII represents an old, differentiated lineage to which one or more large parts of the chromosomal genome of the group IV lineage, including the 40-kb segment on which the invasion genes are located, have been horizontally transferred. All lines of molecular genetic evidence indicate that group V is very strongly differentiated from all other groups, thus supporting its current taxonomic treatment as a species, *Salmonella bongori*, separate from *S. enterica*. The *Salmonella* Reference Collection C, composed of the 16 strains used in DNA sequence studies, has been established for research on variation in natural populations.

Under current taxonomic treatment, strains of the 2,324 serovars of the salmonellae are assigned to seven groups defined on the basis of variation in biotype and the results of a limited number of genomic DNA hybridization experiments (7, 13, 25). In accordance with the long-standing view that all of the salmonellae are members of a single species, designated *Salmonella enterica* (14), the seven groups, numbered I, II, IIIa, IIIb, IV, V, and VI, have generally been considered subspecies (7, 12), but group V was recently elevated to species status, as *S. bongori*, by Reeves et al. (26), a revision endorsed by Brenner (5) and Popoff and Le Minor (25). Groups I, II, and IV correspond to the subgenera of the same designation in Kauffmann's classification (10), and groups IIIa and IIIb include, respectively, the monophasic and diphasic strains of Kauffmann's subgenus III, also known as *Arizona* (7). Group V was recognized in 1982 (16), and group VI was distinguished in 1986 (15).

In the past decade, a considerable volume of data on genetic variation in natural populations of the salmonellae has been accumulating, initially through application of multilocus enzyme electrophoresis (MLEE) (1) and, more recently, by nucleotide sequencing of chromosomal genes encoding enzymes and proteins of several other functional types (32). These molecular genetic analyses have confirmed the division of *Salmonella* into seven distinctive groups and, in addition, have identified an eighth group composed of several strains that were previously assigned to group IV on the basis of biotype.

The objective of this report is to consider the implications of the molecular population genetic data for the taxonomic structure of the salmonellae at the species and subspecies levels. This analysis has provided background for the designation of a reference collection of genetically characterized strains (*Salmonella* Reference Collection C [SARC]) that are representative of the salmonellae as a whole and will be of use to microbiologists and molecular biologists who are concerned with various aspects of genetic and phenotypic variation and evolution in natural populations.

* Corresponding author.

† Present address: Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138.

MATERIALS AND METHODS

Bacterial isolates. This study is based on a collection of 96 strains, including 91 isolates representing all seven of the currently recognized groups of *Salmonella* serovars and 5 isolates that form an eighth group, designated VII. The distribution of the isolates among the groups is as follows: I, 11; II, 21; IIIa, 4; IIIb, 5; IV, 28; V, 13; VI, 9; and VII, 5. The isolates were obtained from the collections of the Centers for Disease Control and Prevention, Atlanta, Ga., and the Institut Pasteur, Paris, France.

SARC. Two representative strains of each of the eight groups were selected to form SARC (Table 1). These strains are the core sample used in a series of comparative DNA sequence analyses in our laboratory, initiated by Nelson et al. (23).

MLEE. All 96 isolates were assayed for electrophoretic variation in the 24 enzymes listed by Selander et al. (28, 29), according to the methods of Selander et al. (30). The five strains of group VII and seven representative strains of group IV were assayed for electrophoretic variation in an additional 11 enzymes: acid phosphatase-3, acid phosphatase-4, alkaline phosphatase-1, alkaline phosphatase-2, citrate synthase, esterase-1, fumarase, glyceraldehyde-3-phosphate dehydrogenase, malic enzyme, glycine-glycine peptidase, and nucleoside phosphorylase-2.

Electromorphs (allozymes) of each enzyme were equated with alleles at the corresponding structural gene locus, and an absence of enzyme activity was attributed to a null allele. Distinctive combinations of alleles (multilocus enzyme genotypes) were designated ETs (electrophoretic types).

Statistical analysis. Genetic diversity at an enzyme locus among ETs was calculated from allele frequencies as $h = (1 - \Sigma x_i^2)(n/n - 1)$, where x_i is the frequency of the ith allele and n is the number of ETs. Mean genetic diversity (H) is the arithmetic average of h over all loci assayed.

For multilocus enzyme data, a neighbor-joining dendrogram (11, 27) was generated from a matrix of Nei's standard genetic distance between pairs of ETs (19) on the basis of the proportion of loci at which dissimilar alleles occur (proportion of mismatches). For nucleotide sequence data, neighbor-joining dendrograms were constructed with the program MEGA (Molecular Evolutionary Genetic Analysis; version 1.0) from matrices of genetic distance between pairs of strains estimated from the number of substitutions per site (9, 20).

Nucleotide sequence accession numbers. The nucleotide sequences of the 12 genes on which dendrograms are based have been deposited in the GenBank

TABLE 1. Properties of the 16 SARC strains

SARC no.	RKS no.	Original no.[a]	Antigenic formula[b]	Source	Locality	Date
Group I						
1	s4194[c]	S 6623	1,4,[5],12:i:1,2	Human	England	1958
2	s3333[d]	IP E.88.374	9,12,[Vi]:d:[Z_{66}]		Dakar	1988
Group II						
3	s2985	CDC 151-85	58:d:z_6	Human	Massachusetts	1985
4	s2993	CDC 3472-64	42:f:g,t:-			1964
Group IIIa						
5	s2980	CDC 346-86	62:z_4,z_{23}:-	Cornsnake	Oregon	1986
6	s2983	CDC 409-85	62:z_{36}:-	Human	California	1985
Group IIIb						
7	s2978	CDC 156-87	$50_{1,2,3}$:k:z	Human	Oregon	1987
8	s2979	CDC 678-94	38[k]:z_{35}:-	Human	California	1984
Group IV						
9	s3015	CDC 2584-68	45a,b:g,z_{51}:-	Animal	Canal Zone	1968
10	s3027	CDC 287-86	16:z_4,z_{32}:-	Human	Illinois	1986
Group V (*S. bongori*)						
11	s3041	CDC 750-72	66:z_{41}:-	Frog		1972
12	s3044	CDC 2703-76	48:z_{41}:-	Parakeet	United States	1976
Group VI						
13	s2995	CDC 1363-65	45:a:e,n,x		India	1965
14	s3057	CDC 347-78	11:b:e,n,x			1978
Group VII						
15	s3013	CDC 2439-64	1,40:g,z_{51}:-		Tonga-T1	1964
16	s3014	CDC 5039-68	40:z_4,z_{24}:-	Human	Florida	1968

[a] Abbreviations: CDC, Centers for Disease Control and Prevention (Atlanta, Ga.); IP, Institut Pasteur (Paris, France).
[b] The antigenic formula indicates the O:H1:H2 factors; see reference 12 for explanation.
[c] Serovar *S. enterica* Typhimurium. The strain was obtained from D. C. Old.
[d] Serovar *S. enterica* Typhi.

database under the following accession numbers: *gapA*, M66853 and M66882; *putP*, L01132 to L01159; *gnd*, U14423 to U14509; *mdh*, U04742 to U04784 and U04786; *aceK*, U43344 to U43359; *spaO*, *spaP*, and *spaO*, U29345 to U29365; *spaM* and *spaN*, U43300 to U43315; and *invA* and *invE*, U43237 to U43274.

RESULTS

Multilocus enzyme genotypes. Among the 96 isolates examined, all 24 enzyme loci were polymorphic, and a comparison of allele profiles identified 80 distinctive ETs. The estimated genetic relationships among the ETs are indicated in the neighbor-joining dendrogram shown in Fig. 1. All but four of the ETs fall into seven discrete clusters that correspond precisely to groups I, II, IIIa, IIIb, IV, V, and VI, as previously defined on the basis of biotype and genomic DNA hybridization (6, 15, 16, 33). However, ETs 56 to 59, represented by five isolates, form a distinctive eighth group (labeled VII).

The isolates of ETs in cluster VII had been identified as members of group IV on the basis of biotype, but in multilocus enzyme genotype they differ from strains of group IV in alleles at 15 of the 24 loci assayed. To better estimate genetic distance, we examined variation at an additional 11 enzyme loci among the five isolates of group VII and single isolates of seven ETs representing the major branches of the group IV cluster. The strains of groups IV and VII differed at 7 of these 11 loci. Thus, in total, ETs of group VII could be distinguished from those of group IV by the occurrence of distinctive alleles at 22 (63%) of the total of 35 loci assayed (Table 2).

In agreement with evidence from genomic DNA hybridization (and multiple gene sequences; see below), MLEE analysis indicated that strains of group V (*S. bongori*) are the most divergent forms of the salmonellae (Fig. 1). But for the other groups (*S. enterica*), a statistical analysis of 1,000 computer-generated dendrograms (bootstrapping) revealed that their evolutionary relationships are not reliably indexed by the topology of the MLEE dendrogram (Fig. 1), the branches of

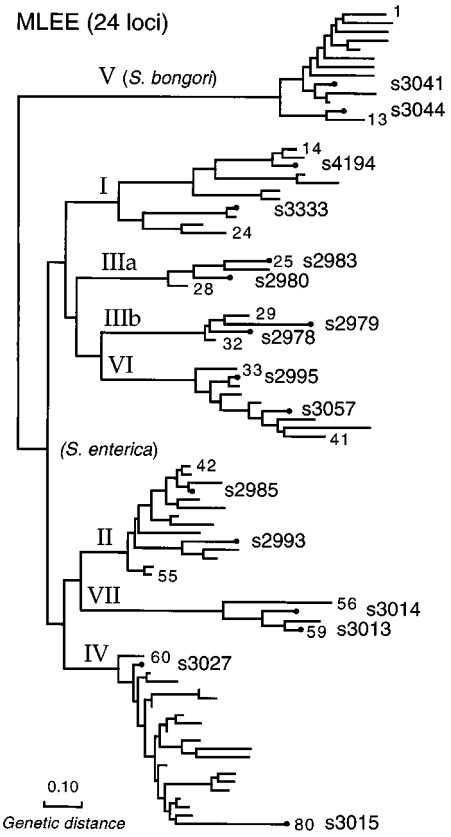

FIG. 1. Neighbor-joining dendrogram for 80 ETs of the eight groups of the genus *Salmonella*, based on the proportion of 24 enzyme loci at which dissimilar alleles (electromorphs) occur (mismatches). The eight groups are designated by roman numerals, and strain numbers (e.g., s3041) are indicated for the 16 strains that constitute SARC.

TABLE 2. Allele profiles for 22 polymorphic enzyme loci in strains of groups IV and VII

Strain	Allele at indicated enzyme locus[a]																					
	AP1	NP1	PLP	IPO	SHK	ES1	CST	AP3	AP4	NP2	G3P	FUM	THD	IDH	ACO	CAK	PGI	LG1	GDH	PGM	GLU	MPI
Group IV																						
s3029	1.9	5.5	1.5	3	7	0	3	3	3	3	1	3	3	2	2.5	3	5	6	3	2	3	4
s3012	1.9	8.2	1.5	3	7	0	3	3	3	3	1	3	3	2	2.5	3	5	6	3	2	3	4
s3037	1.9	5.5	1.5	3	7	0	3	3	3	3	1	3	3	2	2.5	3	5	6	3	2	3	4
s3021	1.9	5.5	2.9	3	7	0	3	3	3	3	1	3	3	3	2.5	3	5	6	3	2	3	4
s3034	1.9	5.5	2.9	3	7	0	3	3	3	3	1	0	3	1	2.5	3	5	6	3	5	3	6
s3015	1.9	5.5	1.5	3	7	0	4	3	0	3	1	3	3	2	0	3	5	6	3	5	3	2.5
s3016	1.9	5	2.9	3	7	0	3	3	3	3	1	3	3	2	2.5	3	5	5	3	2	3	4
Group VII																						
s3013	9	6	8	5	2	3	1	2	1	0	5	5	1	3	3	5	6	2	4	6	2	5
s3014	8	6	8	5	2	3	1	2	2	0	5	5	1	3	3	5	6	3	4	3	2	8
s3801	8	6	8	5	2	3	1	2	2	0	5	5	1	3	3	5	6	3	4	3	2	8
s3803	9	6	8	5	2	3	1	2	1	0	5	5	1	3	3	5	6	2	0	6	2	5
s3804	9	6	7	5	0	0	1	2	0.8	0	4	5	1	3	4	5	5	4	0	6	2	8

[a] Abbreviations: AP1, AP3, and AP4, acid phosphatase-1, -3, and -4; NP1, nucleoside phosphorylase-1; PLP, phenylalanyl-leucine peptidase; IPO, indophenol oxidase; SHK, shikimate dehydrogenase; ES1, esterase-1; CST, citrate synthase; NP2, nucleoside phosphorylase-2; G3P, glyceraldehyde-3-phosphate dehydrogenase; FUM, fumarase; THD, threonine dehydrogenase; IDH, isocitrate dehydrogenase; ACO, aconitase; CAK, carbamylate kinase; PGI, phosphoglucose isomerase; LG1, leucylglycyl-glycine peptidase-1; GDH, glucose dehydrogenase; PGM, phosphoglucomutase; GLU, glutamate dehydrogenase; MPI, mannose phosphate isomerase.

[b] Isolate s3801, which was obtained from the Institut Pasteur (original number IP498/88), may represent the same strain as s3014, which was obtained from the Centers for Disease Control and Prevention (original number CDC 5039-68); they have the same serotype and are identical in multilocus enzyme genotype.

which diverge from one another within a narrow range of genetic distance values. At the relatively deep level of divergence of these groups, most of the enzyme loci assayed are polymorphic for multiple alleles (electromorphs) and the ETs of each group have unique alleles at many loci, with the consequence that the order of branching of the groups is determined by variation at only a few loci. The accuracy of estimates of genetic distance is further limited by the occurrence of multiple amino acid substitutions in individual proteins and the frequent convergence of electromorphs resulting from different amino acid substitutions that produce similar effects on electrophoretic mobility (3, 22).

For the 80 ETs shown in Fig. 1, mean total genetic diversity per locus (H_T), calculated from allele frequencies at the 24 enzyme loci assayed, was 0.627, which may be compared with a value of 0.343 reported (31) for 62 ETs of the ECOR collection of *Escherichia coli* (8, 24). Mean diversity per locus within groups (subsamples) (H_S) was 0.271, and the amount of the total diversity apportioned among groups (G_{ST}), calculated as $(H_T\text{-}H_S)/H_T$, was 0.568, or slightly more than twice that within groups.

Housekeeping gene sequences. For the 16 SARC strains, the nucleotide sequences of five housekeeping genes, glyceraldehyde-3-phosphate dehydrogenase (*gapA*) (23), proline permease (*putP*) (21), malate dehydrogenase (*mdh*) (3), 6-phosphogluconate dehydrogenase (*gnd*) (22), and isocitrate dehydrogenase kinase/phosphatase (*aceK*) (20a), are available. Estimates of the overall genetic relationships among the 16 strains, as indexed by variation in the combined (concatenated) nucleotide sequences of the five genes, are shown in Fig. 2A. The topology of the dendrogram is fully consistent with evidence from DNA hybridization experiments and may, therefore, be considered broadly indicative of the actual evolutionary relationships of the groups, notwithstanding the occasional occurrence of horizontal transfer of gene segments among them (32, 34).

Invasion gene sequences. A 40-kb segment near 59 min on the *S. enterica* chromosome contains 15 or more loci, the *inv*/*spa* genes, whose products are required for the invasion of epithelial cells (18). Sequences of seven of these genes, *invA*, *invE*, *spaM*, *spaN*, *spaO*, *spaP*, and *spaQ*, have been obtained for the SARC strains (2a, 17). The topology of a dendrogram

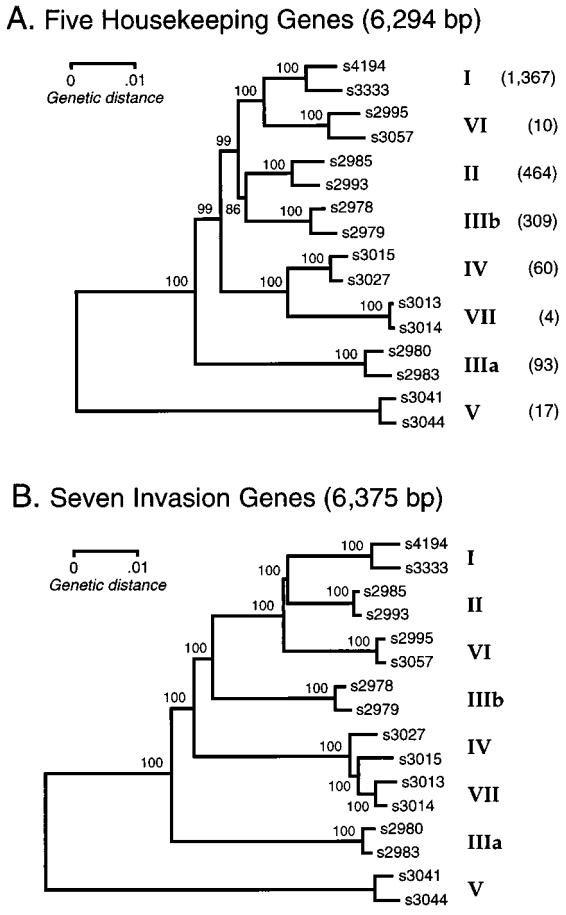

FIG. 2. (A) Neighbor-joining dendrogram for the 16 SARC strains, based on variation in the combined coding sequences of five housekeeping genes. The eight groups are designated by roman numerals, and the number of serotypes assigned to each group is indicated in parentheses. Bootstrap values based on 1,000 computer-generated trees are indicated at the nodes. (B) Neighbor-joining trees for seven invasion genes, based on variation in the combined coding sequences.

constructed from their combined sequences (Fig. 2B) is similar to the comparable dendrogram for the five housekeeping genes (Fig. 2A), except for the position of group II relative to groups IIIb and VI and, surprisingly, an absence of substantial differentiation between strains of groups IV and VII.

DISCUSSION

Species limits. On the basis of an MLEE analysis and the reported evidence from genomic DNA hybridization experiments, Reeves et al. (26) designated the group V salmonellae as a species (*S. bongori*) distinct from the other groups (*S. enterica*). Support for this taxonomic interpretation is provided by our MLEE study and a comparative nucleotide sequence analysis of 12 housekeeping and invasion genes, both of which have clearly demonstrated that the group V strains are strongly differentiated from all other salmonellae.

Evolutionary status of group VII. This group presents a paradoxical situation. For each of the five housekeeping genes sequenced, the degree of nucleotide sequence divergence between group VII and group IV is at least as great as that between groups I and VI or II and IIIb (Fig. 2A) and by MLEE analysis of 35 enzymes, group VII is strongly differentiated from group IV (Table 2) and all other groups (Fig. 1). Yet there is little differentiation between VII and IV in the sequences of any of the seven invasion genes studied. Thus, it appears that the genome of the strains of group VII is a mosaic of large chromosomal segments with different evolutionary histories. As an explanatory hypothesis, we suggest that the major part of the chromosomal genome of the group VII strains is that of a old lineage that had achieved, through mutation and selection, a level of genetic differentiation equivalent to that of the other groups of *S. enterica*. The five housekeeping genes that we have sequenced and most or all of the enzyme loci assayed by MLEE presumably are located on this part of the chromosome. Subsequently, one or more chromosomal segments, including the 40-kb segment which contains the *inv/spa* genes, were horizontally transferred from the lineage of group IV to that of group VII. Consistent with this interpretation are the results of a genomic DNA hybridization experiment in which the reassociation values obtained in a comparison of two strains of VII with three strains of IV were intermediate between those reported for similar comparisons of the other groups of *S. enterica* and those normally obtained for strains of the same group (24a).

Reference collections. All 94 strains on which this study was based, including the 16 strains that constitute SARC, have been deposited in the *Salmonella* Genetic Stock Centre (contact K. E. Sanderson, Department of Biological Sciences, University of Calgary, 200 University Drive NW, Calgary, Alberta, Canada T2N 1N4). SARC is the third reference collection of strains established for the salmonellae. SARA (2) consists of 72 strains of several serovars of the *S. enterica* Typhimurium complex, and SARB (4) includes 72 strains representing 37 serovars of group I.

ACKNOWLEDGMENTS

We thank J. J. Farmer III, L. Le Minor, and M. Y. Popoff for providing strains. J. Li provided sequence data for *spaM* and *spaN*, and K. Nelson furnished sequences of *aceK* and generated the dendrogram for the housekeeping genes.

This research was supported by grant AI22144 from the National Institutes of Health.

REFERENCES

1. **Beltran, P., J. M. Musser, R. Helmuth, J. J. Farmer III, W. M. Frerichs, I. K. Wachsmuth, K. Ferris, A. C. McWhorter, J. G. Wells, A. Cravioto, and R. K. Selander.** 1988. Toward a population genetic analysis of *Salmonella*: genetic diversity and relationships among strains of serotypes *S. choleraesuis*, *S. derby*, *S. dublin*, *S. enteritidis*, *S. heidelberg*, *S. infantis*, *S. newport*, and *S. typhimurium*. Proc. Natl. Acad. Sci. USA **85:**7753–7757.
2. **Beltran, P., S. A. Plock, N. H. Smith, T. S. Whittam, D. C. Old, and R. K. Selander.** 1991. Reference collection of strains of the *Salmonella typhimurium* complex from natural populations. J. Gen. Microbiol. **137:**601–606.

2a. **Boyd, E. F., and J. Li.** Unpublished data.

3. **Boyd, E. F., K. Nelson, F.-S. Wang, T. S. Whittam, and R. K. Selander.** 1994. Molecular genetic basis of allelic polymorphism in malate dehydrogenase (*mdh*) in natural populations of *Escherichia coli* and *Salmonella enterica*. Proc. Natl. Acad. Sci. USA **91:**1280–1284.
4. **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.
5. **Brenner, D. J.** 1991. Introduction to the family Enterobacteriaceae, p. 2673–26952. *In* H. Balow, H. G. Trüper, M. Dworkin, W. Harder, and K.-H. Schleifer (ed.), The prokaryotes, 2nd ed., vol. 3. Springer-Verlag, New York.
6. **Crosa, J. H., D. J. Brenner, W. H. Ewing, and S. Falkow.** 1973. Molecular relationships among the salmonellae. J. Bacteriol. **115:**307–315.
7. **Ewing, W. H.** 1986. Edwards and Ewing's identification of Enterobacteriaceae, 4th ed. Elsevier, New York.
8. **Herzer, P. J., S. Inouye, M. Inouye, and T. S. Whittam.** 1990. Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. J. Bacteriol. **172:**6175–6181.
9. **Jukes, T. H., and C. R. Cantor.** 1969. Evolution of protein molecules, p. 21–132. *In* H. N. Munro (ed.), Mammalian protein metabolism. Academic Press, New York.
10. **Kauffmann, F.** 1966. The bacteriology of Enterobacteriaceae. Williams & Wilkins, Baltimore.
11. **Kim, J., F. J. Rohlf, and R. R. Sokal.** 1993. The accuracy of phylogenetic estimation using the neighbor-joining method. Evolution **47:**471–486.
12. **Le Minor, L.** 1984. *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 & Wilkins, Baltimore.
13. **Le Minor, L.** 1991. The genus *Salmonella*, p. 2760–2774. *In* H. Balow, H. G. Trüper, M. Dworkin, W. Harder, and K.-H. Schleifer (ed.), The prokaryotes, 2nd ed., vol. 3. Springer-Verlag, New York.
14. **Le Minor, L., and M. Y. Popoff.** 1987. Designation of *Salmonella enterica* sp. nov., nom. rev., as the type and only species of the genus *Salmonella*. Int. J. Syst. Bacteriol. **37:**465–468.
15. **Le Minor, L., M. Y. Popoff, B. Laurent, and D. Hermant.** 1986. Individualisation d'une septième sous-espèce de *Salmonella*: *S. choleraesuis* subsp. *indica* subsp. nov. Ann. Microbiol. (Inst. Pasteur) **137B:**211–217.
16. **Le Minor, L. M. Véron, and M. Y. Popoff.** 1982. Taxonomie des *Salmonella*. Ann. Microbiol. (Inst. Pasteur) **133B:**223–243.
17. **Li, J., H. Ochman, E. A. Groisman, E. F. Boyd, F. Solomon, K. Nelson, and R. K. Selander.** 1995. Relationship between evolutionary rate and cellular location among the Inv/Spa invasion proteins of *Salmonella enterica*. Proc. Natl. Acad. Sci. USA **92:**7252–7256.
18. **Mills, D. M., V. Bajaj, and C. A. Lee.** 1995. A 40 kb chromosomal fragment encoding *Salmonella typhimurium* invasion genes is absent from the corresponding region of the *Escherichia coli* K-12 chromosome. Mol. Microbiol. **15:**749–759.
19. **Nei, M.** 1987. Molecular evolutionary genetics. Columbia University Press, New York.
20. **Nei, M., and T. Gojobori.** 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:**418–426.

20a. **Nelson, K.** Unpublished data.

21. **Nelson, K., and R. K. Selander.** 1992. Evolutionary genetics of the proline permease gene (*putP*) and the control region of the proline utilization operon in populations of *Salmonella* and *Escherichia coli*. J. Bacteriol. **174:** 6886–6895.
22. **Nelson, K., and R. K. Selander.** 1994. Intergeneric transfer and recombination of the 6-phosphogluconate dehydrogenase gene (*gnd*) in enteric bacteria. Proc. Natl. Acad. Sci. USA **91:**10227–10231.
23. **Nelson, K., T. S. Whittam, and R. K. Selander.** 1991. Nucleotide polymorphism and evolution in the glyceraldehyde-3-phosphate dehydrogenase gene (*gapA*) in natural populations of *Salmonella* and *Escherichia coli*. Proc. Natl. Acad. Sci. USA **88:**6667–6671.
24. **Ochman, H., and R. K. Selander.** 1984. Standard reference strains of *Escherichia coli* from natural populations. J. Bacteriol. **157:**690–693.

24a. **Popoff, M. Y.** Personal communication.

25. **Popoff, M. Y., and L. Le Minor.** 1992. Antigenic formulas of the *Salmonella* serovars. WHO Collaborating Center for Reference and Research on *Salmonella*, Institut Pasteur, Paris.
26. **Reeves, M. W., G. M. Evins, A. A. Heiba, B. D. Plikaytis, and J. J. Farmer III.** 1989. Clonal nature of *Salmonella typhi* and its genetic relatedness to other salmonellae as shown by multilocus enzyme electrophoresis, and proposal of *Salmonella bongori* comb. nov. J. Clin. Microbiol. **27:**311–320.
27. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method

for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.

28. **Selander, R. K., P. Beltran, N. H. Smith, R. M. Barker, P. B. Crichton, D. C. Old, J. M. Musser, and T. S. Whittam.** 1990. Genetic population structure, clonal phylogeny, and pathogenicity of *Salmonella paratyphi* B. Infect. Immun. **58:**1891–1901.
29. **Selander, R. K., P. Beltran, N. H. Smith, R. Helmuth, F. A. Rubin, D. J. Kopecko, K. Ferris, B. D. Tall, A. Cravioto, and J. M. Musser.** 1990. Evolutionary genetic relationships of clones of *Salmonella* serovars that cause human typhoid and other enteric fevers. Infect. Immun. **58:**2262–2275.
30. **Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. N. Gilmour, and T. S. Whittam.** 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl. Environ. Microbiol. **51:** 873–884.
31. **Selander, R. K., D. A. Caugant, and T. S. Whittam.** 1987. Genetic structure and variation in natural populations of *Escherichia coli*, p. 1625–1648. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
32. **Selander, R. K., J. Li, E. F. Boyd, F.-S. Wang, and K. Nelson.** 1994. DNA sequence analysis of the genetic structure of populations of *Salmonella enterica* and *Escherichia coli*, p. 17–49. *In* F. G. Priest, A. Ramos-Cormenzana, and B. J. Tindall (ed.), Bacterial diversity and systematics. Plenum Press, New York.
33. **Stoleru, G. H., L. Le Minor, and A. M. Lhéritier.** 1976. Polynucleotide sequence divergence among strains of *Salmonella* sub-genus IV and closely related organisms. Ann. Microbiol. (Inst. Pasteur) **127A:**477–486.
34. **Thampapapillai, G., R. Lan, and P. R. Reeves.** 1994. Molecular evolution in the *gnd* locus of *Salmonella enterica*. Mol. Biol. Evol. **11:**813–828.