Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 1280–1284, February 1994
Genetics

Molecular genetic basis of allelic polymorphism in malate dehydrogenase (*mdh*) in natural populations of *Escherichia coli* and *Salmonella enterica*

(PCR/substitution rates/NAD$^+$-binding and catalytic domains/multilocus enzyme electrophoresis/clonal population structure)

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

Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park, PA 16802

Contributed by Robert K. Selander, October 20, 1993

ABSTRACT **Nucleotide sequences of the *mdh* gene encoding the metabolic enzyme malate dehydrogenase (MDH) were determined for 44 strains representing the major lineages of *Escherichia coli* and the eight subspecies of *Salmonella enterica*. Sequence diversity was four times greater in *S. enterica* than in *E. coli*, and in both species the rate of amino acid substitution was lower in the NAD$^+$-binding domain than in the catalytic domain. Divergence of the *mdh* genes of the two species apparently has not involved excess nonsynonymous substitutions resulting from the fixation of adaptive amino acid mutations. Allozyme analysis detected 57% of the distinctive amino acid sequences. Statistical tests of the distribution of polymorphic synonymous nucleotide sites identified four possible intragenic recombination events, one involving a single allele of *E. coli* and three involving alleles of three subspecies of *S. enterica*. But recombination at *mdh* has not occurred with sufficient frequency to obscure the phylogenetic relationships among strains indicated by multilocus enzyme electrophoresis, total DNA hybridization, and sequence analysis of the *gapA* and *putP* genes. These findings provide further evidence that the effective (realized) rates of horizontal transfer and recombination for metabolic enzyme and other housekeeping genes are generally low in these species, in contrast to those for loci encoding or mediating the structure of cell-surface and other macromolecules for which recombinants may be subject to strong balancing, directional, or diversifying selection.**

By providing full resolution of allelic diversity in genes, comparative nucleotide sequencing has opened a new era for population genetics, and the availability of a rapidly increasing body of sequence data has stimulated the development of new statistical methods for data analysis and hypothesis testing in studies of the genetic structure of natural populations and the evolutionary processes that affect rates of nucleotide and amino acid substitution. In application to bacteria, sequence analysis has already implicated recombination of horizontally transferred DNA as a factor in the evolution of one or more genes in several diverse species and has further demonstrated marked variation in allelic diversity and effective (realized) rate of recombination among loci. For example, in both *Escherichia coli* and *Salmonella enterica*, recombination at the *gnd* locus encoding 6-phosphogluconate dehydrogenase has occurred with sufficient frequency to greatly increase allelic diversity (1); in the case of *E. coli*, it has all but obscured the phylogenetic relationships among strains (refs. 2–6; K.N. and R.K.S., unpublished data). But, in contrast, the sequences of several other metabolic enzyme genes, including glyceraldehyde-3-phosphate dehydrogenase (*gapA*), and the sequence encoding the transport protein proline permease (*putP*) show little or no evidence that recombination has contributed to either allelic or multilocus genotypic diversity (7–9).

We report here the results of a comparative analysis of sequence variation in the polymorphic malate dehydrogenase gene (*mdh*) among 44 strains of *E. coli* and *S. enterica* recovered from natural populations.† The molecular genetic basis of allozyme variation in the malate dehydrogenase (MDH) protein is revealed, and relative rates of sequence divergence within and between species are estimated. And our findings provide further evidence that effective recombination rates in these enteric species are generally low for metabolic enzyme genes.

MATERIALS AND METHODS

Bacterial Isolates. Most of the strains on which this study is based were previously examined for sequence variation in *gapA* (7) and *putP* (8). From the *E. coli* Reference Collection (ECOR) (10) and the T.S.W. research collection, we chose 19 strains representing five major evolutionary lineages that have been identified by multilocus enzyme electrophoresis (MLEE) (11, 12) and one strain (E830587) that has not been assigned to an ECOR group.

A sample of 24 *S. enterica* strains representing all eight recognized subspecies (13) was selected from the *Salmonella* Reference Collections (14).

Enzyme Electrophoresis. Cell lysates were electrophoresed and stained for MDH activity (15), and allozymes (electromorphs) were designated alphabetically in order of decreasing anodal mobility.

PCR Amplification and Nucleotide Sequencing. From each strain of *E. coli*, a 864-bp segment (codons 12–299) of the 939-bp coding region of the *mdh* gene was PCR-amplified from extracted total DNA. Primers for PCR were designed from the published sequences of laboratory strain K-12 (16, 17), as follows: 5′ primer (5′-ATGAAAGTCGCAGTCCTCGGCGCTGCTGGCGG-3′) and 3′ primer (5′-TTAACGAACTCCTGCCCCAGAGCGATATCTTTCTT-3′).

For amplification of *mdh* from strains of *S. enterica*, a 3′ primer (5′-ATATCTTTYTTCAGCGTATCCAGCAT-3′) was designed from the published sequence of laboratory strain LT2 (18). Use of this primer in conjunction with the *E. coli* 5′ primer yielded an 849-bp segment (codons 12–294) of the coding region.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MDH, malate dehydrogenase; MLEE, multilocus enzyme electrophoresis; ECOR, *E. coli* Reference Collection.

*To whom reprint requests should be addressed.

†The sequences reported in this paper have been deposited in the GenBank data base (accession nos. U04742–U04784, U04786).

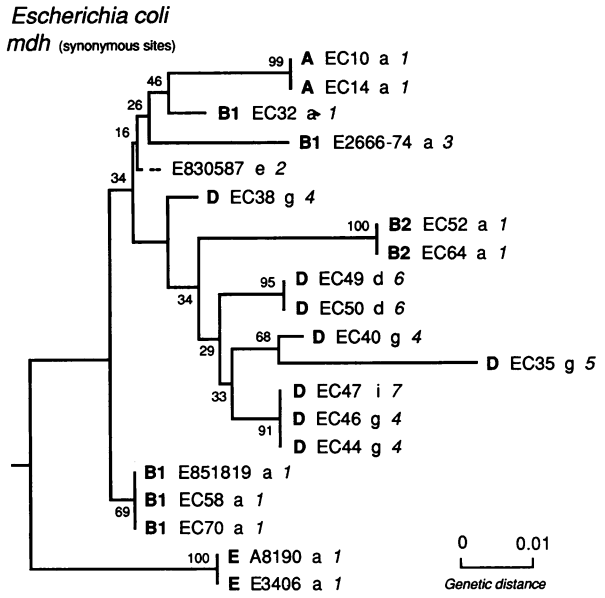

FIG. 1. Evolutionary tree for the *mdh* gene in 20 strains of *E. coli*, constructed by the neighbor-joining method (20) from a matrix of pairwise distances based on synonymous nucleotide sites. Strain numbers, ECOR group assignments (uppercase letters; ref. 13), MDH allozymes (lowercase letters), and amino acid sequence numbers are indicated (see Fig. 3). A number adjacent to a node indicates the percentage of 1000 bootstrap trees that contained the node.

Single-stranded DNA was generated by the λ exonuclease procedure (19), and the resulting template was sequenced in both orientations (Sequenase; United States Biochemical).

RESULTS

Sequence Variation. Among the 20 *mdh* sequences of *E. coli* examined, there were 40 polymorphic nucleotide sites, and pairs of strains differed, on average, at 1.1% of nucleotide sites and 0.3% of amino acid positions. There were 13 distinctive nucleotide sequences and six sets of identical sequences (see Fig. 1).

Variation among the 24 *S. enterica* sequences was four times greater than that among the *E. coli* sequences. There were 142 polymorphic nucleotide sites, and pairs of strains differed, on average, at 4.5% of nucleotide sites and 1.3% of amino acid positions, with most of the variation being attributable to differences between subspecies. The nucleotide sequences of the two strains of subspecies VII were identical, as were those of two of the three strains of subspecies II (Fig. 2).

There were 217 polymorphic nucleotide sites among all 44 sequences, and strains of *S. enterica* and *E. coli* differed at an average of 124 sites (14.6%). In all, 30 amino acid positions were polymorphic, with an average species difference of 12.4 (4.3%), which includes 9 positions at which there are fixed differences between the species.

Distribution of Synonymous Polymorphic Sites. To detect possible cases of intragenic recombination, we used the Stephens (21) test of nonrandom clustering of polymorphic nucleotide sites. For the 33 polymorphic synonymous sites among the 20 *E. coli* sequences, the only significant phylogenetic partition separated the sequence of strain EC35 from the remaining sequences and was supported by five clustered sites (278, 308, 662, 707, and 710) ($g_o = 354$, $P = 0.022$). This significant clustering was also detected by the Sawyer (3) test (sum of squared condensed fragment lengths = 46,942, $P \leq 0.015$).

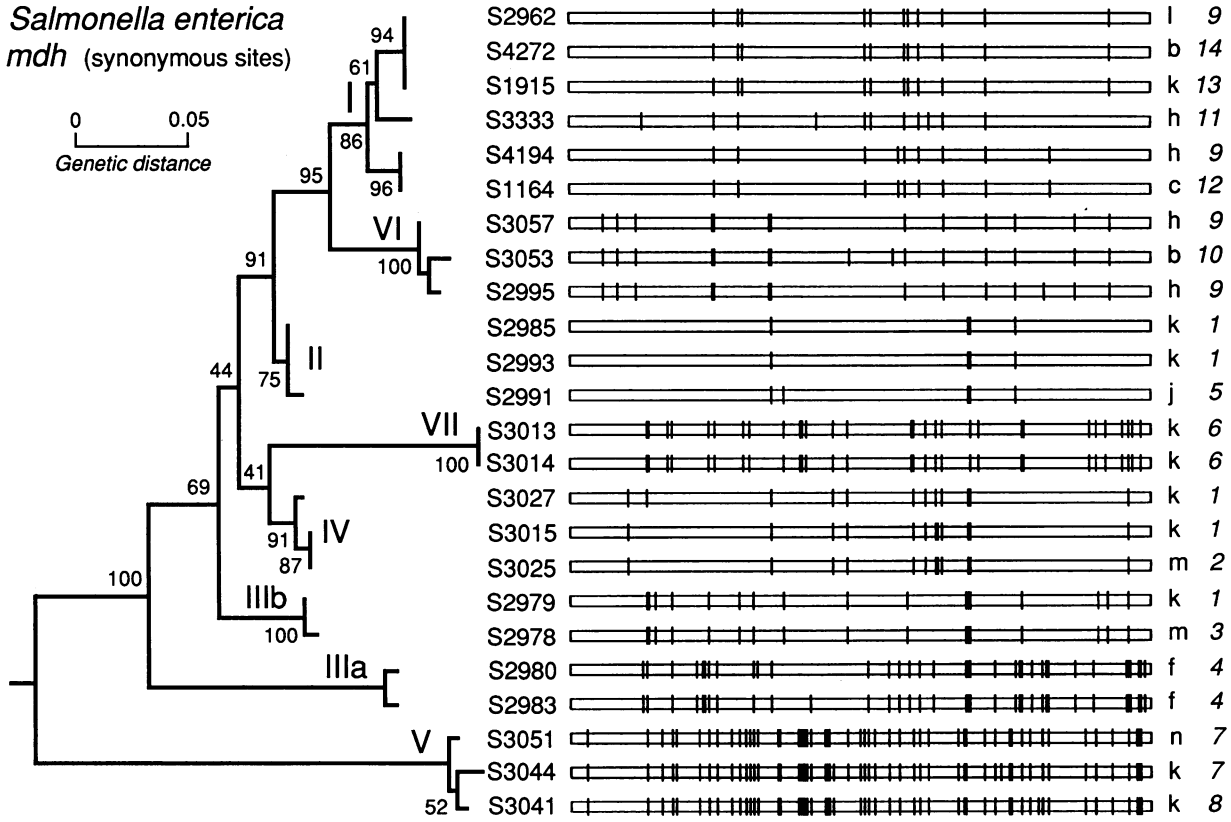

FIG. 2. Evolutionary tree for the *mdh* gene in 24 strains of *S. enterica*, based on synonymous nucleotide sites. The locations along the gene of synonymous substitutions from the consensus sequence are indicated by vertical lines. Subspecies clusters are indicated by Roman numerals; strain numbers (center), MDH allozymes (lowercase letters), and amino acid sequence numbers (at right) are indicated (see Figs. 4 and 5).

Among the 24 *S. enterica* sequences, there were 115 polymorphic synonymous sites, the positions of which are shown in Fig. 2. The Stephens test identified 49 phylogenetic partitions, 34 of which were trivial, involving single sites. Of the remaining 15 partitions, three were supported by multiple synonymous sites that were nonrandomly clustered. The first significant partition separated the two sequences of subspecies IIIa from all other sequences and was supported by a total of 12 sites, including 3 sites at the 5′ end of the gene (194, 203, and 206) and 9 sites at the 3′ end (542, 668, 686, 707, 776, 824, 830, 851, and 852). For a pattern of randomly placed differences, the stretch of 336 nt between these two clusters of polymorphic sites is highly improbable (g_o = 336, $P \leq 0.01$). The second partition grouped the three sequences of subspecies VI and was supported by a cluster of 5 sites (59, 80, 107, 218, and 302) (d_o = 243, $P \leq 0.05$). The third partition, which separated the five sequences of subspecies V and VII from the remaining sequences, was supported by only 2 sites (350 and 356), but the probability of the occurrence of 2 polymorphic sites as close as or closer than 6 nt is 0.014. Note that this partition grouped the subspecies IIIa sequences with those of subspecies I, II, IIIb, IV, and VI and that the two sites are in the stretch that defines the first partition. Application of the Sawyer test failed to detect overall clustering of sites.

Rates of Substitution in Functional Domains. Estimates of the numbers (per 100 sites) of synonymous and nonsynonymous nucleotide substitutions (d_S and d_N) (22, 23) were obtained separately for regions of the *mdh* gene encoding the NAD^+-binding domain (amino acids 1–150) and the catalytic domain (amino acids 151–313) of MDH (24). In both *E. coli* and *S. enterica*, nonsynonymous substitutions have occurred less frequently in the binding domain than in the catalytic domain. For *E. coli*, values of d_N for the binding domain, the catalytic domain, and both domains combined were 0.09 ± 0.09, 0.20 ± 0.15, and 0.15 ± 0.09, respectively. Comparable d_N values for *S. enterica* were 0.03 ± 0.03, 1.13 ± 0.35, and 0.58 ± 0.18. In both *E. coli* and *S. enterica*, d_S was also smaller for the segment of the gene encoding the binding domain than for that specifying the catalytic domain, although for neither species is the difference statistically significant. For the binding domain, the catalytic domain, and both domains combined, d_S values were as follows: *E. coli*, 1.88 ± 0.65, 5.53 ± 1.18, and 3.70 ± 0.65; *S. enterica*, 14.09 ± 2.00, 23.35 ± 2.65, and 18.53 ± 1.61.

Evolutionary Trees for *mdh* Sequences. Relationships among strains indicated by synonymous sites of the *mdh* sequences of the *E. coli* strains (Fig. 1) are similar to those previously determined by MLEE (12) and sequence analysis of *gapA* (7) and *putP* (8). ECOR group A strains are closely related in all four data sets, as are group B2 strains; and group D strains occur in a single cluster with the B2 strains. The nucleotide sequences of all three genes have indicated only a weak association among strains of ECOR group B1.

In the *mdh* tree for *S. enterica* (Fig. 2), the genes cluster by subspecies, and the branching order is similar to that indicated by *gapA* and *putP* trees and by MLEE (7, 8). The only distinctive feature of the topology of the *mdh* tree is that subspecies I is more similar to subspecies VI than to subspecies II, whereas subspecies I and II cluster together in both the *gapA* and *putP* trees. This difference suggests a recombination event, presumably that detected by the nonrandom clustering of polymorphic synonymous sites in the 5′ region of the *mdh* gene.

Amino Acid Substitutions Underlying Allozyme Variation. Electrophoresis identified 14 distinctive allozymes of MDH, 5 in *E. coli* and 9 in *S. enterica*. None was shared by the two species, and the sequences of their commonest allozymes differed at 10 amino acid positions.

For *E. coli*, the relationships between allozymes and amino acid sequences are shown in Fig. 1, and a hypothetical evolutionary scheme is presented in Fig. 3. Of the seven distinctive amino acid sequences, sequence 1 is regarded as ancestral because it occurs in 10 strains representing all ECOR groups except D. The scenario (Fig. 3) postulates six amino acid replacements at five codon positions; each of four of these replacements involved a charge change and generated a new allozyme. One parallel change is indicated: from the common ancestor of ECOR group D strains, sequence 6 (allozyme d) was derived by an Asp-71 → Asn substitution before the occurrence of a Gln-289 → Lys substitution that produced sequence 4 (allozyme g); subsequently, a second independent Asp-71 → Asn replacement created sequence 7 (allozyme i).

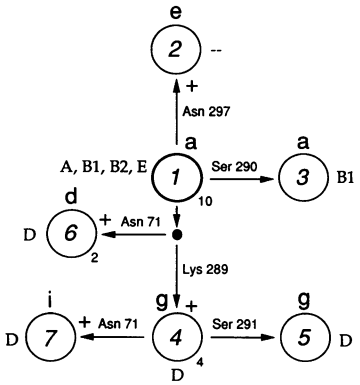

FIG. 3. Hypothetical evolutionary scheme for the allozymes (lowercase letters) and amino acid sequences (circled numbers) of strains of *E. coli*. Uppercase letters indicate ECOR group assignments of strains having the various sequences. For sequences represented by more than one strain, the number of strains is indicated to the lower right of the circle.

Codon number

Amino acid sequence	63	160	165	172	191	195	201	240	246	247	258	276	279	281	286	290	293	294
1	Ala GCA	Glu GAG	Leu CTG	Pro CCG	Gly GGC	Thr ACC	Asp GAT	Arg CGC	Lys AAA	Gly GGC	Gly GGT	Glu GAA	Lys AAA	Ile ATC	Ala GCT	Arg CGC	Glu GAA	Gly GGT
2	Ala ...	.C.	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3	...	...	...	...	...	...	...	...	...	..T	...	Gly .G.	...	...	...	...	...	...
4	...	...	...	...	...	...	...	...	Thr .C.	..G	...	...	..G	...	...	...	...	..C
5	...	...	...	...	...	...	...	His .A.	...	...	Ser A..	...	...	...	...	...	...	...
6	...	...	...	...	...	...	...	...	...	...	...	...	Arg .G.	Val G.T	...	...	...	...
7	...	...	...	...	...	Ser .G.	Glu ..A	..T	...	...	..C	...	...	..T	...	...	...	...
8	...	...	...	Ser T..	...	Ser .G.	Glu ..A	..T	...	...	..C	...	...	..T	...	...	...	...
9	...	...	...	...	...	...	Glu ..G	...	...	...	...	...	...	...	...	His .AT	Asp ..C	Ala .C.
	...	...	...	...	...	...	..G	...	..G	...	...	...	...	...	...	.AT	..C	.C.
	...	...	T..	...	...	...	..G	...	...	...	...	...	...	...	...	.AT	..T	.C.
10	...	...	T..	...	Asp .A.	...	Glu ..G	...	...	...	...	...	...	...	...	His .AT	Asp ..T	Ala .C.
11	...	...	...	...	...	...	Glu ..G	...	...	...	...	...	...	...	Thr A..	His .AT	Asp ..C	Ala .C.
12	...	...	...	...	...	...	Glu ..G	Leu .T.	..G	...	...	...	...	...	...	His .AT	Asp ..C	Ala .C.
13	...	Arg .G.	...	...	...	...	Glu ..G	...	...	...	...	...	...	...	...	His .AT	Asp ..C	Ala .C.
14	Val .T.	...	...	...	...	...	Glu ..G	...	...	Asp .A.	...	...	...	...	...	His .AT	Asp ..C	Ala .C.

FIG. 4. MDH amino acid sequences of strains of *S. enterica*. All codons in which one or more nucleotide substitutions result in amino acid replacements are shown. Dots indicate nucleotide identity with sequence number 1.

An evolutionary scheme for the nine allozymes and 14 amino acid sequences (Fig. 4) of *S. enterica* is shown in Fig. 5. In addition to the 14 observed amino acid sequences, we assume the existence of 2 sequences (designated as 7*x* and 9*x* in Fig. 5) that have one or more amino acid substitutions in the segment at the 3′ end of the gene (codons 283–313) that was not sequenced. Thus, the scenario postulates 15 evolutionary events, but the nature of only 13 of them is known from the available sequence data.

From amino acid sequence 1 (allozyme k), allozymes f, j, and m were derived by single or double substitutions, with allozyme m arising twice (sequences 2 and 3). Sequence 9 (allozyme h), which is regarded as ancestral for strains of subspecies I and VI, was derived from sequence 1 (allozyme k) by four amino acid substitutions. Subsequently, four single and one double amino acid replacements generated allozymes b and c, with b arising twice, once by an Asp-191 substitution and again by Val-63 and Asp-247 changes. Another case of convergence involved an Arg-165 substitution, which produced an allozyme variant (sequence 13) that is indistinguishable from allozyme k (sequences 1, 6, 7, and 8).

In sum, all the allozyme variation in *E. coli* can be accounted for by single amino acid substitutions that resulted in unit positive increases in charge. In *S. enterica*, both positive and negative charge changes have occurred; and, additionally, two allozymes (h and j) were generated by replacements of amino acids of like charge.

DISCUSSION

Amino Acid Sequence Basis of Allozyme Variation. Comparative nucleotide sequence analysis of 90–92% of the *mdh* gene has revealed the amino acid sequence diversity underlying allozyme variation in MDH in 20 strains of *E. coli* and 24 strains of *S. enterica*. In these 44 strains, protein electrophoresis detected 14 allozymes. A total of 21 distinctive amino acid sequences was identified, but we postulate the existence of two additional *S. enterica* sequence variants involving replacement nucleotide substitutions in the unsequenced 3′ end of the gene. Hence, 12 (57%) of the 21 observed amino acid sequences were recognized on the basis of allozyme variation. Our evolutionary scenarios for the alleles of the two species postulated the occurrence of 19 evolutionary steps, involving a total of 26 amino acid substitutions. Thus, on average, the generation of a new allozyme has involved 2.6 amino acid substitutions.

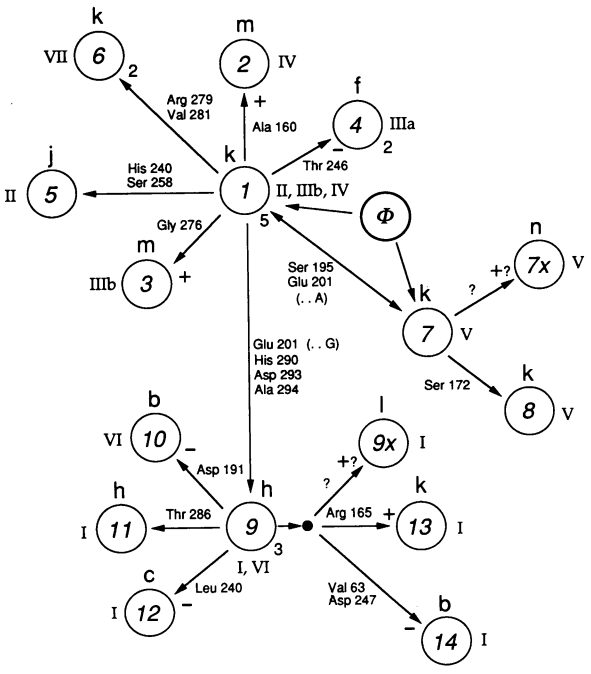

FIG. 5. Hypothetical evolutionary scheme for the allozymes (lowercase letters) and amino acid sequences (circled numbers) of strains of *S. enterica*. Roman numerals indicate subspecies assignments of strains having the various sequences. Φ designates an unobserved hypothetical ancestral sequence. See Fig. 3 and text for further explanation.

In *S. enterica*, three cases of the production of the same allozyme by different amino acid substitutions were identified. In each of two of these cases (allozymes b and m), the convergent allozymes were derived from the same ancestral sequence by single- or double-amino acid substitutions. But in the case of allozyme k, proteins differing by as many as eight amino acids (sequences 8 and 13) were electrophoretically indistinguishable.

Divergence Between Species. Our analysis identified nine codons specifying fixed amino acid differences between *E. coli* and *S. enterica*. To test the prediction of the neutral theory that the ratios of nonsynonymous to synonymous substitutions are equal between species and between alleles within species, we used the method of Whittam and Nei (25). The number of synonymous substitutions per 100 synonymous sites estimated to have occurred since the time of divergence of *E. coli* and *S. enterica* from a common ancestor (d_A) is 66.91 ± 2.06, and the average within-species level of diversity (π) is 11.11 ± 0.03. Comparable values for nonsynonymous substitutions are $d_A = 2.06 \pm 0.21$ and $\pi = 0.338 \pm 0.001$. The ratio of nonsynonymous to synonymous substitutions between species is 0.031 ± 0.011, and the ratio of the within-species diversities is 0.030 ± 0.016. Similar results were obtained by application of the method of analysis developed by McDonald and Kreitman (26). Thus, there is no evidence that the divergence of the *mdh* sequences of *E. coli* and *S. enterica* has involved excess nonsynonymous substitutions resulting from the fixation of adaptive amino acid mutations.

Genetic Structure of Populations. The concept of clonal population structure has had major impact on the fields of bacterial epidemiology and pathogenesis (1, 27–31) and has recently been applied to certain parasitic protozoans as well (32). Indeed, it is widely regarded as a paradigm for all bacteria (33–35), although we have noted that linkage disequilibrium analyses of MLEE data do not support a clonal structure for the pathogenic species *Neisseria gonorrhoeae* and *Pseudomonas aeruginosa* (28) or major phylogenetic divisions of the soil bacterium *Rhizobium meliloti* (36). It is unlikely that any bacterial species will prove to be strictly clonal, and certainly none is panmictic or consists of "promiscuous, freely intermixing populations" (34). The real problem is to determine the contribution of intragenic and assortative recombination to allelic and genotypic variation in diverse types of bacteria by estimating the frequency and extent of exchange for a large variety of genes.

The analysis of the *mdh* locus reported here, taken together with comparable data for *gapA* and *putP* in *E. coli* and *S. enterica* (7, 8) and several other enzyme loci in *E. coli* (9), strongly suggests that the effective rate of recombination is low for most metabolic enzyme and other "housekeeping" genes. [Recombination of small segments of DNA at a low rate in the *trp* region in *E. coli* has also been reported (37).] Thus, the findings for these loci are consistent with the results of MLEE analysis and other lines of evidence indicating that the population structure of *E. coli* and *S. enterica* is basically clonal in the sense that effective recombination rates for most genes are low enough to permit the mutational diversification of cell lineages in terms of host distribution, disease specificity, and virulence and the maintenance of differentially adapted chromosomal genotypes over periods sufficiently

long for them to become widely distributed geographically (28, 38).

For the highly polymorphic flagellin *fliC* locus in *S. enterica*, we have shown that horizontal transfer and recombination are a major source of allelic and serovar diversity (refs. 6 and 39; J. Li and R.K.S., unpublished data). Because flagellin is a highly antigenic cell-surface protein that interacts directly with the external environment, recombinants may have an immediate adaptive advantage (e.g., by presenting altered cell-surface structures to host defense mechanisms) and be brought to high frequency in local populations by natural selection and then transferred to other lineages (6, 8, 39). Similar explanations in terms of adaptation may apply to a number of genes in other bacteria for which evidence of horizontal transfer and recombination is available, including those encoding pili in *E. coli* (40), enzymes of the restriction modification system in *E. coli* (41), the class 1 outer membrane protein in *Neisseria meningitidis* (42), the M1 protein in group A *Streptococcus* (43), and penicillin-binding proteins in *Neisseria* spp. (44) and *Streptococcus pneumoniae* (45). But for housekeeping genes, such as *mdh*, it seems unlikely that either intragenic or assortative recombination would result in a selective advantage to the recipient cell. And the probable fate of deleterious or selectively neutral recombinants is loss from the population through purifying selection and genetic drift.

In sum, the evidence suggests that differing modes and strengths of natural selection among loci encoding or mediating the synthesis of products of different functional type influence the rate of effective recombination in populations. Although the 6-phosphogluconate dehydrogenase gene (*gnd*) is a conspicuous exception to the generalization that the effective rate of recombination is low for metabolic enzyme genes, it is one that may prove the rule. The *gnd* locus lies near the genes of the *rfb* region, which mediate synthesis of the highly polymorphic cell-surface lipopolysaccharide. There is reason to believe that these genes are subject to strong frequency-dependent selection for the production of lipopolysaccharide antigenic diversity (29), and it has been suggested that the proximity of *gnd* to the polymorphic lipopolysaccharide genes diminishes the chance of recombinant *gnd* alleles being lost by genetic drift (4, 5). Several apparent cases of the horizontal cotransfer of a polymorphic cell-surface lipopolysaccharide serotype and a *gnd* allozyme between strains of *E. coli* have been reported (11, 31).

In conclusion, the lesson emerging from comparative sequence analysis of bacterial genes is that recombination rates and other evolutionary processes cannot be determined for entire genomes from data for single loci. And extrapolation from one bacterial species to another or even between different populations of the same species may yield erroneous conclusions. For example, whereas populations of *Neisseria meningitidis* serogroup B and C show only weak linkage disequilibrium and thus may be considered only short-term or "epidemically" clonal (35), those of serogroup A exhibit strong disequilibrium and are among the more highly clonal bacterial populations yet studied (30).

We thank S. A. Plock and S. Spigelmeyer for technical assistance and S. W. Schaeffer for helpful discussion. This research was supported by Public Health Service Grants AI-22144, AI-24566, and AI-00964 from the National Institutes of Health.

1. Whittam, T. S. & Ake, S. E. (1993) in *Mechanisms of Molecular Evolution*, eds. Takahata, N. & Clark, A. G. (Sinauer, Sunderland, MA), pp. 223–245.
2. Barcak, G. J. & Wolf, R. E., Jr. (1988) *J. Bacteriol.* **170**, 372–379.
3. Sawyer, S. (1989) *Mol. Biol. Evol.* **6**, 526–538.
4. Biserčić, M., Feutrier, J. Y. & Reeves, P. R. (1991) *J. Bacteriol.* **173**, 3894–3900.
5. Dykhuizen, D. E. & Green, L. (1991) *J. Bacteriol.* **173**, 7257–7268.
6. Selander, R. K., Nelson, K., Li, J., Boyd, E. F., Wang, F.-S. & Whittam, T. S. (1994) in *Identification of Bacteria: Current Status and Future Prospects*, eds. Priest, F. G., Ramos-Cormenza, A. & Tindall, B. (Plenum, New York), in press.
7. Nelson, K., Whittam, T. S. & Selander, R. K. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 6667–6671.
8. Nelson, K. & Selander, R. K. (1992) *J. Bacteriol.* **174**, 6886–6895.
9. Hall, B. G. & Sharp, P. M. (1992) *Mol. Biol. Evol.* **9**, 654–665.
10. Ochman, H. & Selander, R. K. (1984) *J. Bacteriol.* **157**, 690–693.
11. Selander, R. K., Caugant, D. A. & Whittam, T. S. (1987) in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, eds. Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M. & Umbarger, H. E. (Am. Soc. Microbiol., Washington, DC), Vol. 2, pp. 1625–1648.
12. Herzer, P. J., Inouye, S., Inouye, M. & Whittam, T. S. (1990) *J. Bacteriol.* **172**, 6175–6181.
13. Selander, R. K., Beltran, P. & Smith, N. H. (1991) in *Evolution at the Molecular Level*, eds. Selander, R. K., Clark, A. G. & Whittam, T. S. (Sinauer, Sunderland, MA), pp. 25–57.
14. Boyd, E. F., Wang, F.-S., Beltran, P., Plock, S. A., Nelson, K. & Selander, R. K. (1993) *J. Gen. Microbiol.* **139**, 1125–1132.
15. Selander, R. K., Caugant, D. A., Ochman, H., Musser, J. M., Gilmour, M. N. & Whittam, T. S. (1986) *Appl. Environ. Microbiol.* **51**, 873–884.
16. McAlister-Henn, L., Blaber, M., Bradshaw, R. A. & Nisco, S. J. (1987) *Nucleic Acids Res.* **15**, 4993.
17. Vogel, R. F., Entian, K.-D. & Mecke, D. (1987) *Arch. Microbiol.* **149**, 36–42.
18. Lu, C.-D. & Abdelal, A. T. (1993) *Gene* **123**, 143–144.
19. Higuchi, R. G. & Ochman, H. (1989) *Nucleic Acids Res.* **17**, 5865.
20. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
21. Stephens, J. C. (1985) *Mol. Biol. Evol.* **2**, 539–556.
22. Nei, M. & Gojobori, T. (1986) *Mol. Biol. Evol.* **3**, 418–426.
23. Nei, M. & Jin, L. (1989) *Mol. Biol. Evol.* **6**, 290–300.
24. Hall, M. D., Levitt, D. G. & Banaszak, L. J. (1992) *J. Mol. Biol.* **226**, 867–882.
25. Whittam, T. S. & Nei, M. (1991) *Nature (London)* **354**, 115–116.
26. McDonald, J. H. & Kreitman, M. (1991) *Nature (London)* **351**, 652–654.
27. Caugant, D. A., Frøholm, L. O., Bøvre, K., Holten, E., Frasch, C. E., Mocca, L. F., Zollinger, W. D. & Selander, R. K. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 4927–4931.
28. Selander, R. K. & Musser, J. M. (1990) in *Molecular Basis of Bacterial Pathogenesis*, eds. Iglewski, B. H. & Clark, V. L. (Academic, San Diego), pp. 11–36.
29. Reeves, P. R. (1992) *FEMS Microbiol. Lett.* **100**, 509–516.
30. Wang, J.-F., Caugant, D. A., Li, X., Hu, X., Poolman, J. T., Crowe, B. A. & Achtman, M. (1992) *Infect. Immun.* **60**, 5267–5282.
31. Whittam, T. S., Wolfe, M. L., Wachsmuth, I. K., Ørskov, F., Ørskov, I. & Wilson, R. A. (1993) *Infect. Immun.* **61**, 1619–1629.
32. Tibayrenc, M., Neubauer, K., Barnabé, C., Guerrini, F., Skarecky, D. & Ayala, F. J. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 1335–1339.
33. Istock, C. A., Duncan, K. E., Ferguson, N. & Zhou, X. (1992) *Mol. Ecol.* **1**, 95–103.
34. Lenski, R. E. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 4334–4336.
35. Maynard Smith, J., Smith, N. H., O'Rourke, M. & Spratt, B. G. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 4384–4388.
36. Eardly, B. D., Materon, L. A., Smith, N. H., Rumbaugh, M. D. & Selander, R. K. (1990) *Appl. Environ. Microbiol.* **56**, 187–194.
37. Milkman, R. & Bridges, M. M. (1993) *Genetics* **133**, 455–468.
38. Achtman, M. & Pluschke, G. (1986) *Annu. Rev. Microbiol.* **40**, 185–210.
39. Smith, N. H. & Selander, R. K. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 956–960.
40. Marklund, B.-I., Tennent, J. M., Garcia, E., Hamers, A., Båga, M., Lindberg, F., Gaastra, W. & Normark, S. (1992) *Mol. Microbiol.* **6**, 2225–2242.
41. Murray, N. E., Daniel, A. S., Cowan, G. M. & Sharp, P. M. (1993) *Mol. Microbiol.* **9**, 133–143.
42. Feavers, I. M., Heath, A. B., Bygraves, J. A. & Maiden, M. C. J. (1992) *Mol. Microbiol.* **6**, 489–495.
43. Harbaugh, M. P., Podbielski, A., Hugl, S. & Cleary, P. P. (1993) *Mol. Microbiol.* **8**, 981–991.
44. Spratt, B. G., Bowler, L. D., Zhang, Q.-Y., Zhou, J. & Maynard Smith, J. (1992) *J. Mol. Evol.* **34**, 115–125.
45. Laibile, G., Spratt, B. G. & Hakenbeck, R. (1991) *Mol. Microbiol.* **5**, 1993–2002.