Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 10227–10231, October 1994
Genetics

Intergeneric transfer and recombination of the 6-phosphogluconate dehydrogenase gene (*gnd*) in enteric bacteria

(*Escherichia coli*/*Salmonella enterica*/*Citrobacter*/*Klebsiella*/O antigen polysaccharide)

KIMBERLYN NELSON AND ROBERT K. SELANDER

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

Contributed by Robert K. Selander, July 5, 1994

ABSTRACT **The *gnd* gene, encoding 6-phosphogluconate dehydrogenase (EC 1.1.1.44), was sequenced in 87 strains of 15 species assigned to five nominal genera of the Enterobacteriaceae, including 36 isolates of *Salmonella enterica* and 32 strains of *Escherichia coli*. In *S. enterica*, the effective (realized) rate of recombination of horizontally transferred *gnd* sequences is only moderately higher than the rates for other chromosomal housekeeping genes. In contrast, recombination at *gnd* has occurred with such high frequency in *Escherichia coli* that the indicated evolutionary relationships among strains are not congruent with those estimated by sequence analysis of other genes and by multilocus enzyme electrophoresis. *E. coli* and *S. enterica* apparently have not exchanged *gnd* sequences, but those of several strains of *E. coli* have been imported from species of *Citrobacter* and *Klebsiella*. The relatively frequent exchange of *gnd* within and among taxonomic groups of the Enterobacteriaceae, compared with other housekeeping genes, apparently results from its close linkage with genes that are subject to diversifying selection, including those of the *rfb* region determining the structure of the O antigen polysaccharide.**

The *gnd* gene encodes 6-phosphogluconate dehydrogenase (6PGD; EC 1.1.1.44), the third enzyme of the pentose phosphate pathway, which is one of two central and constitutive routes of intermediary carbohydrate metabolism in enteric bacteria, glycolysis being the other. Because the enzyme has an important metabolic function, the expectation from evolutionary theory is that both the amino acid sequence of 6PGD and the nucleotide sequence of *gnd* should be relatively conserved. But in populations of *Escherichia coli*, allozyme (electromorph) diversity of 6PGD is three times greater than that of the average enzyme of similar molecular size (1), and several studies of nucleotide sequence variation in *gnd* have identified interstrain transfer and recombination as a factor contributing to an unusually high level of allelic (genic) diversity (2–4).

6PGD is also highly polymorphic in populations of *Salmonella enterica* (5–7), and the number of allozyme alleles is larger than expected from estimates of genic diversity (Fig. 1), which suggests that sequence variation in *gnd* in this species has also been generated by recombination. To examine this possibility and to determine the extent to which *gnd* sequences have been transferred between *E. coli* and *S. enterica* and among related types of bacteria, we have sequenced *gnd* in 87 strains of species belonging to five nominal genera of the family Enterobacteriaceae, including 36 strains of *S. enterica* and 32 strains of *E. coli*.* Our analysis revealed that recombination of *gnd* sequences has occurred with lesser frequency in *S. enterica* than in *E. coli*,

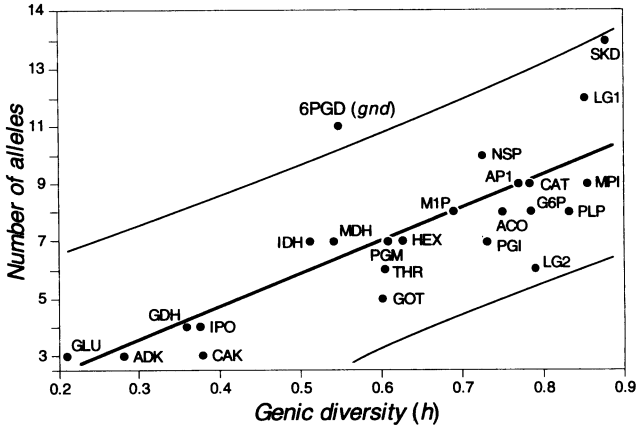

FIG. 1. Relationship between number of alleles and genic diversity for 24 enzyme loci among 80 electrophoretic types of the eight subspecies of *S. enterica* (8). Enzyme abbreviations are defined in ref. 8. Thin lines above and below the regression line indicate the 95% confidence limits for individual predicted values.

in which genic diversity has been augmented by recruitment of alleles from species of *Klebsiella* and *Citrobacter*.

MATERIALS AND METHODS

Bacterial Strains. This study is based on 87 strains of enteric bacteria representing 15 species of *Escherichia, Shigella, Salmonella, Citrobacter*, and *Klebsiella*. Sixteen of the 36 strains of *S. enterica* (two each of the eight subspecies) and 13 of the 32 strains of *E. coli* in the sample were previously examined for sequence variation in several other genes (9–11). Twenty other isolates of *S. enterica* were selected from the *Salmonella* Reference Collections (7) on the basis of 6PGD allozyme variation, and 19 additional strains of *E. coli* were obtained from the *E. coli* Reference Collection (ECOR) (12, 13).

In addition to these samples, 19 strains of 13 species of other genera were examined, as follows: six strains of *Citrobacter diversus* and one of *Citrobacter amalonaticus* (14), *Citrobacter freundii* (ATCC 8090), *Escherichia fergusonii* (ATCC 35469), *Escherichia vulneris* (ATCC 33821), two strains of *Klebsiella pneumoniae* (ATCC 13883, C. H. Whitfield), *Klebsiella planticola* (ATCC 33531), *Klebsiella terrigena* (ATCC 33257), *Klebsiella* sp. (C. H. Whitfield), *Shigella boydii* (ATCC 8700), *Shigella dysenteriae* (ATCC 13313), *Shigella flexneri* (ATCC 29903), and *Shigella sonnei* (ATCC 29930).

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: 6PGD, 6-phosphogluconate dehydrogenase; ECOR, *E. coli* reference collection; MLEE, multilocus enzyme electrophoresis.

*The sequences discussed in this paper have been deposited in the GenBank data base (accession nos. U14423–U14509).

Nucleotide Sequencing. DNA isolation, PCR amplification, nucleotide sequencing, and assembly and editing of sequences followed previously described protocols (15). For each strain, 1335 bp of the 1404-bp *gnd* gene were sequenced on both strands; and the sequences were aligned with the Eyeball Sequence Editor (16). Primers for PCR were designed from the published sequences of *S. enterica* serovar Typhimurium strain LT2 (17) and *E. coli* strain K-12 (18).

Evolutionary Trees. Molecular evolutionary relationships among *gnd* sequences were examined by the neighbor-joining method of tree construction (19), based on pairwise genetic distances estimated from numbers of synonymous substitutions (20, 21). The significance of the branching order was evaluated by bootstrap analysis of 1000 computer-generated trees. All analyses were performed with MEGA (22).

RESULTS

Variation Within Species. Average values of sequence divergence in *gnd* and 6PGD within *S. enterica* and *E. coli* are presented in Table 1. For pairs of strains of *S. enterica*, nucleotide sequence divergence (3.82%) was within the range (3.8–4.6%) previously estimated for *gapA* (glyceraldehyde-3-phosphate dehydrogenase) (9), *mdh* (malate dehydrogenase) (11), and *putP* (proline permease) (10), but there was a lesser degree of amino acid divergence in 6PGD (0.72%) than in the proteins encoded by these other genes. In contrast, both nucleotide and amino acid sequence divergences among strains of *E. coli* were more than twice as large as those obtained for *gapA*, *mdh*, *putP*, and *aceK* (isocitrate dehydrogenase phosphatase/kinase) (K.N., unpublished data). And the maximum observed interstrain *gnd* sequence divergence in the *E. coli* sample (19.3%) was larger than comparable values previously reported (3, 4).

Amino Acid Substitutions and Allozymes. In an analysis of *mdh* variation in *S. enterica* and *E. coli*, 57% of the amino acid sequences were detected as distinctive allozymes, pairs of which differed, on average, by 2.6 amino acid substitutions (11). For 6PGD, a similar level of resolution of allelic variation was obtained for *S. enterica* and *C. diversus*, but not for *E. coli*. In *S. enterica*, 15 (58%) of 26 distinctive 6PGD amino acid sequences were detected by enzyme electrophoresis, and allozyme pairs differed, on average, by 4.2 amino acids. However, some allozymes were highly heterogeneous in sequence; for example, among 15 strains with the most common allozyme, there were 10 distinctive sequences differing from one another by as many as seven amino acids. In *C. diversus*, three allozymes and four amino acid sequences were identified.

In *E. coli*, only 10 (38%) of 26 amino acid sequences were detected electrophoretically. Many of the 6PGD allozymes were markedly heterogeneous in amino acid sequence, as previously reported (3), but some were all but uniform. For example, strains EC49, EC50, and EC57 uniquely shared the same allozyme and had similar amino acid sequences, although they belong to three different ECOR groups and are distinctive in the sequences of other genes examined.

Evolutionary Relationships Within *Salmonella*. Evolutionary relationships among the *gnd* alleles of *S. enterica* (Fig. 2) are generally similar to the relationships among the strains indicated by DNA hybridization (23, 24), multilocus enzyme electrophoresis (MLEE), and the sequences of other genes (8–11). Subspecies V sequences are the most divergent, followed by those of subspecies IIIa; and except for subspecies IV and VI, the sequences of strains of the same subspecies are closely similar.

Table 1. Average pairwise sequence divergence among 36 strains of *S. enterica* and 32 strains of *E. coli*

	gnd nucleotides		6PGD amino acids	
Species	No. of differences	%	No. of differences	%
S. enterica	51.5	3.82	3.2	0.72
E. coli	95.7	7.15	8.0	1.79

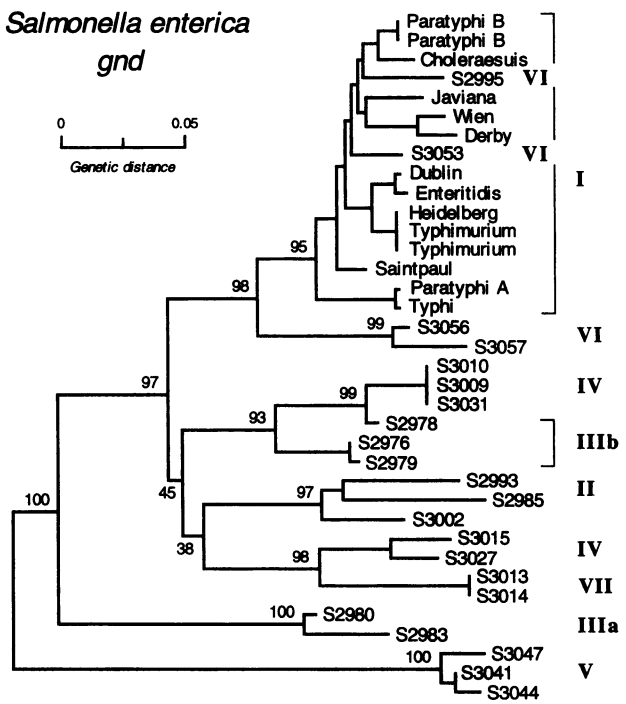

FIG. 2. Neighbor-joining tree for *gnd* sequences of strains of *S. enterica*, based on synonymous sites. Roman numerals are subspecies designations. Bootstrap values are indicated at the nodes.

All 14 subspecies I serovars cluster together on the basis of *gnd* sequence, but two of the four subspecies VI strains also occur within the subspecies I group. The isolates of subspecies IV are also divided into two groups; two strains are in a cluster allied with that of subspecies VII (which is the relationship typically evidenced by other genes), while the other three strains are grouped with isolates of subspecies IIIb. The fact that the *gnd* sequences of strains of subspecies II are far removed from those of subspecies I and VI (Fig. 2) is also anomalous, since all other lines of molecular evidence indicate relatively close relationships among these three subspecies (8).

Within the subspecies I cluster, strains of serovars that are known to be closely related tend to be associated, as, for example, Heidelberg and Typhimurium (5, 25) and Dublin and Enteritidis (5, 26). However, the close similarity between Typhi and Paratyphi A is at odds with evidence that strains of these serovars differ in allozyme alleles at 50% of their enzyme loci (6). Their similarity in *gnd* sequence is largely attributable to a sharing of a cluster of seven unique polymorphic sites in a 210-bp segment ($P = 0.0009$) (27).

Evolutionary Relationships Within *E. coli*. Recombination at the *gnd* locus—both intragenic and assortative (entire gene) (1)—has been rampant in *E. coli*. As reported from other studies of *gnd* polymorphism in this species (3, 4), recombination has been so extensive that a tree based on *gnd* (Fig. 3) bears little resemblance to an MLEE tree (13) or to trees for other genes (9–11). Strains of each of the five major ECOR groups (A, B1, B2, D, and E) are widely distributed over the *gnd* tree, which has a bush-like topology, reflecting frequent intragenic exchange.

Intergeneric Gene Transfer. Evolutionary relationships among the *gnd* sequences of all 87 enteric bacteria are shown

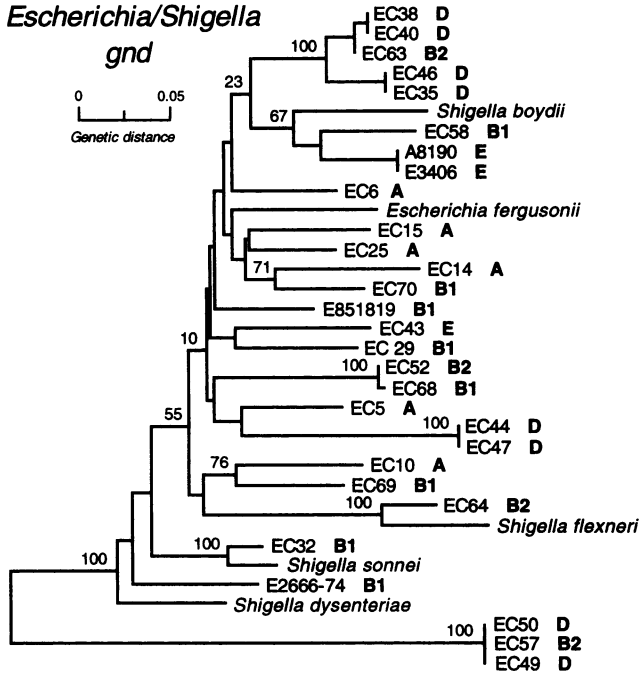

FIG. 3. Neighbor-joining tree for *gnd* sequences of strains of *E. coli* and *Shigella* spp., based on synonymous sites. ECOR group affinities of strains are indicated in boldface. EC16, EC71, and E830587 are not shown (see Fig. 4). Bootstrap values are indicated at the nodes.

in Fig. 4. With the exception of the distribution of several strains of *E. coli* (see beyond), the relationships depicted are not very different from those that might be expected from the findings of other molecular studies, although at first sight certain aspects of the phylogeny appear to be unusual. That

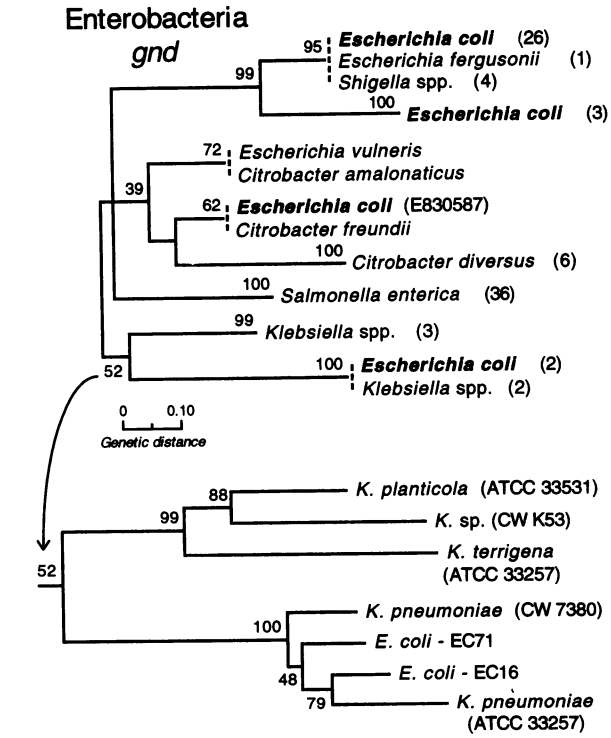

FIG. 4. Neighbor-joining tree for *gnd* sequences of 87 enteric bacteria, based on synonymous sites. Numbers of strains sequenced are indicated in parentheses. Bootstrap values are indicated at the nodes.

the *gnd* alleles of the four *Shigella* species occur within the major group of *E. coli* sequences (Figs. 3 and 4) is not a surprising result and need not be attributed to intergeneric transfer, because DNA hybridization, MLEE, and other lines of evidence clearly indicate that strains of *Shigella* and *E. coli* are sufficiently similar to be considered conspecific (28–30). From an evolutionary genetic standpoint, the shigellae are pathogenic, human-adapted clonal lineages of *E. coli* (31, 32).

Inasmuch as recent DNA hybridization studies have demonstrated that *Citrobacter* is a highly heterogeneous assemblage of strains (33), the divergence in *gnd* sequence among species of *Citrobacter* apparently also reflects a taxonomic problem rather than a phenomenon of special interest to evolutionary geneticists. MLEE data (13), DNA hybridization results (33), and the sequences of *gapA* (K.N., unpublished data) indicate that *C. amalonaticus* is genetically very different from *C. diversus* and *C. freundii*. And an association of *E. vulneris* with *C. amalonaticus* is also evident in the sequences of *gapA* (34). Because studies of several other genes have also indicated that certain species presently assigned to *Escherichia* are distantly related (34), the position of *E. vulneris* relative to *E. coli* in the *gnd* tree is not surprising; and, conversely, *E. fergusonii* is closely related to *E. coli* in *gapA* and *ompA* sequences (34), as well as those of *gnd*.

But artifacts of taxonomy and nomenclature aside, our study has yielded clear evidence of several cases of intergeneric transfer of *gnd*. The sequences of *E. coli* strains EC16 and EC71 are, on average, 94% similar to those of *Klebsiella pneumoniae* and only 84% similar to those of strains in the main *E. coli* group (Fig. 4). These two sequences fall within a group of five *Klebsiella* sequences that includes those of *K. pneumoniae*, *K. planticola*, and *K. terrigena*. In contrast, the sequences of *gapA*, *putP*, and *aceK* clearly place EC16 and EC71 within the larger body of *E. coli* isolates, and MLEE readily distinguishes them from *Klebsiella*.

E. coli strain E830587 clusters with *C. freundii* (Fig. 4), although other gene sequences and MLEE data indicate that in overall genomic character it is well within the main body of *E. coli*. Finally, three strains (EC49, EC50, and EC57) diverge at an average of 9% of nucleotides from the other 26 strains in the main *E. coli* group.

DISCUSSION

Genetic Linkage and Hitchhiking. Because it is most unlikely that the unusually high effective rate of recombination in *gnd* can be attributed to the direct action of diversifying selection on the locus itself, another explanation must be sought. The most plausible one is its close linkage to a number of loci that determine the structure of cell-surface macromolecules, including genes of the *rfb* region that mediate biosynthesis of the highly antigenic polysaccharide domain (O antigen) of the somatic lipopolysaccharide (1, 3, 4, 35). Genes of the *rfb* region are believed to be subject to strong frequency-dependent selection in connection with the avoidance of host defense systems (36); there is evidence that recombination occurs frequently in the *rfb* region and that some or all of these genes in *S. enterica* and *E. coli* have been recruited from other, distantly related (and unidentified) types of bacteria (32, 36).

Recent studies of the segment of the *E. coli* and *S. enterica* chromosomes in which *gnd* is located (44–45 minutes) have revealed the following gene order: the *cps* family (determining capsular polysaccharide structure), *galF*, the *rfb* region, *gnd*, *rol*, and the *his* operon (37, 38). The distance from the end of the *rfb* region to *gnd* varies from 200 to 2000 bp; there are several open reading frames between *gnd* and the *his* operon, including the *rol* (or *cld*) locus, which encodes a product regulating modal chain length of the O antigen (37,

38); this locus is strain specific and may be important in interactions with host cell membranes and immune systems. The inference is that the frequency of recombination at *gnd* is influenced by the action of selection for allelic diversity at neighboring loci, as *gnd* sequences hitchhike with adaptive recombinants of *rfb* and, perhaps, other genes. The anomalous position of the three *S. enterica* subspecies IV strains S3010, S3009, and S3031 (Fig. 2), which are all but identical in *gnd* sequence and share O antigen 50, may be explained by cotransfer of *gnd* and *rfb* from a strain of subspecies IIIb such as S2978, which also expresses O50. In view of our finding of almost identical *gnd* sequences in the serovars Typhi and Paratyphi A, which show no particular relationship on the basis of MLEE (6) or sequences of other housekeeping genes, it is noteworthy that the sequences of their *rfb* genes are also closely similar (36), clearly indicating cotransfer. A possible mechanism for the dissemination of O factors among strains was recently identified by the demonstration that isolates of *S. enterica* serovar Borreze carry a plasmid that bears *rfb* genes encoding O antigen 54 (39). For strains of *E. coli*, several cases have been identified in which alleles determining an O antigen serotype and a 6PGD allozyme apparently have been horizontally cotransferred (40, 41).

The most likely explanation for the similarity of the *gnd* alleles of *E. coli* strains EC16 and EC71 and those of *Klebsiella* is that these sequences were horizontally transferred from *Klebsiella*. The structure of the O-specific polysaccharide is identical in some serotypes of *E. coli* and *Klebsiella* (42), and recent studies of *Klebsiella* O3 and *E. coli* O9 strains have demonstrated considerable sequence similarity throughout the *rfb* region (43). An analysis of the 3′ end of the *E. coli* O9 *rfb* region (44) revealed that the direction of transcription of *gnd* and several *rfb* genes is the reverse of the usual one and that the *gnd* sequence of O9 is most similar to that of EC16, which we have shown to be like that of *Klebsiella*. By the use of PCR primers located in *gnd* and *his*, we have determined (unpublished data) that the gene arrangement in EC16, EC71, and *K. pneumoniae* is like that of *E. coli* O9.

The sequences of the *rol* gene in *E. coli* antigen types O75 and O111 reportedly are as different from one another (66% sequence similarity) as each is from those of *S. enterica* serovar Typhimurium (37, 38). This circumstance and the relatively low codon adaptation index and G+C content of these genes strongly suggests horizontal transfer. Preliminary analysis of the segment extending from *gnd* to *his* indicates that its history in *E. coli* is similar to that of *gnd* and may have involved the transfer of the entire region within and among genera (K.N., unpublished data).

Other Examples of Intergeneric Transfer. On the basis of G+C content, Ochman and Lawrence (45) concluded that perhaps 10% of the genes of *E. coli* and *S. enterica* have been acquired by horizontal transfer from outside sources. However, more direct evidence of intergeneric transfer from sequence analysis is available for only a few genes. Among different strains of *E. coli*, genes of the type I restriction-modification systems may be as divergent in sequence as they are from homologous sequences in *S. enterica*. And there is evidence that part of the *hsdM* gene has been transferred between *S. enterica* and *E. coli* (46). The inference is that highly divergent *hsd* sequences acquired by recombination may be retained in populations as a result of phage-mediated selection.

Sequences of genes encoding two GDP-mannose pyrophosphorylases (*rfbM* and *cpsB*) and two phosphomannomutases (*rfbK* and *cpsG*) in *S. enterica* and *E. coli* show little relation to species boundaries, which may reflect the retention of ancient allele families in the two genera, recent episodes of horizontal gene transfer, or both (44).

In an analysis of five supposedly *Salmonella*-specific cloned DNA fragments (47), Groisman *et al.* (48) identified one occurring on a plasmid in *Klebsiella pneumoniae* and another in certain isolates of *Shigella* and *E. coli*, in some of which the sequence was chromosomal in location. Another possible case of intergeneric transfer has been reported for the *argF* genes of *Klebsiella* and *E. coli* (49).

Finally, the *viaB* region, which consists of genes involved in expression of the virulence polysaccharide capsule (Vi antigen) (50, 51), has clearly been subject to intergeneric transfer, since it is known to occur only in *C. freundii* and in *S. enterica*, in which it is confined to serovars Typhi, Paratyphi A, Paratyphi C, and a geographically localized subclone of Dublin (6, 52).

Interlocus Variation in Recombination Rate. That genes encoding metabolic enzymes and other types of housekeeping proteins are likely to have relative low effective rates of recombination has now been empirically demonstrated for *S. enterica* and *E. coli* from sequence data for at least nine genes widely distributed on the chromosome (8). This finding is consistent with evidence of strong linkage disequilibrium from MLEE studies (53, 54) and several other types of information indicating that the population structure of *S. enterica* and *E. coli* is basically clonal in the sense that the effective recombination rates for most genes are low enough to permit the mutational diversification of cell lineages in terms of biochemical characteristics and ecological niche relationships, including host distribution, disease specificity, and virulence, and the long-term, if not permanent, maintenance of differentially adapted, widely distributed chromosomal genotypes in populations (40, 55–58). In these species and in many other bacteria, genes that are known to experience unusually frequent horizontal transfer and recombination almost without exception encode or mediate the expression of products for which there would seem to be a premium on structural diversity per se or which confer adaptive traits such as antibiotic resistance (8). Thus, for example, intragenic recombination within and among subspecies of *S. enterica* is a major evolutionary mechanism generating allelic diversity in the *fliC* gene, which encodes the highly antigenic and polymorphic flagellar protein; and assortative recombination is a primary source of new serovars (59).

Because it encodes a typical metabolic enzyme, the *gnd* locus is an obvious exception to this generalization, particularly in *E. coli*. However, as we have noted, *gnd* is located within a segment of the chromosome where selection for allelic diversity in several neighboring loci (*cps*, *rfb*, and *rol*) may be strong, with the consequence that the resulting increases in effective rates of recombination at these loci have secondarily elevated the frequency of recombination in *gnd*. This postulated effect has resulted in the near randomization of *gnd* sequences among strains of *E. coli*, but in *S. enterica* it has produced only a moderate increase in the frequency of recombination compared with that of other metabolic enzymes. This difference may be attributed in large part, if not entirely, to ecological differences between the two species, as these determine opportunities for genetic exchange within species and with other enterobacteria. As a basically commensal species, *E. coli*, like *Klebsiella* spp. and *Citrobacter* spp., is a common element of the rich intestinal flora of higher vertebrates (60), whereas *S. enterica* is a specialized intracellular pathogen.

We thank C. H. Whitfield for providing *Klebsiella* strains. This research was supported by Grant AI-22144 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. Biserčić, M., Feutrier, J. Y. & Reeves, P. R. (1991) *J. Bacteriol.* **173**, 3894–3900.
4. Dykhuizen, D. E. & Green, L. (1991) *J. Bacteriol.* **173**, 7257–7268.
5. Beltran, P., Musser, J. M., Helmuth, R., Farmer, J. J., III, Frerichs, W. M., Wachsmuth, I. K., Ferris, K., McWhorter, A. C., Wells, J. G., Cravioto, A. & Selander, R. K. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7753–7757.
6. Selander, R. K., Beltran, P., Smith, N. H., Helmuth, R., Rubin, F. A., Kopecko, D. J., Ferris, K., Tall, B. D., Cravioto, A. & Musser, J. M. (1990) *Infect. Immun.* **58**, 2262–2275.
7. Boyd, E. F., Wang, F.-S., Beltran, P., Plock, S. A., Nelson, K. & Selander, R. K. (1993) *J. Gen. Microbiol.* **139**, 1125–1132.
8. Selander, R. K., Li, J., Boyd, E. F., Wang, F.-S. & Nelson, K. (1994) in *Bacterial Systematics and Diversity*, eds. Priest, F. G., Ramos-Cormenzana, A. & Tindall, R. (Plenum, New York), in press.
9. Nelson, K., Whittam, T. S. & Selander, R. K. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 6667–6671.
10. Nelson, K. & Selander, R. K. (1992) *J. Bacteriol.* **174**, 6886–6895.
11. Boyd, E. F., Nelson, K., Wang, F.-S., Whittam, T. S. & Selander, R. K. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 1280–1284.
12. Ochman, H. & Selander, R. K. (1984) *J. Bacteriol.* **157**, 690–693.
13. Herzer, P. J., Inouye, S., Inouye, M. & Whittam, T. S. (1990) *J. Bacteriol.* **172**, 6175–6181.
14. Li, J., Musser, J. M., Beltran, P., Kline, M. W. & Selander, R. K. (1990) *J. Clin. Microbiol.* **28**, 1760–1765.
15. Nelson, K. & Selander, R. K. (1994) *Methods Enzymol.* **235**, 174–183.
16. Cabot, E. L. & Beckenbach, A. T. (1989) *Comp. Appl. Biosci.* **5**, 233–234.
17. Reeves, P. & Stevenson, G. (1989) *Mol. Gen. Genet.* **217**, 182–184.
18. Nasoff, M. S., Baker, H. V., II, & Wolfe, R. E., Jr. (1984) *Gene* **27**, 253–264.
19. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
20. Nei, M. & Gojobori, T. (1986) *Mol. Biol. Evol.* **3**, 418–426.
21. Jukes, T. H. & Cantor, C. R. (1969) in *Mammalian Protein Metabolism*, ed. Munro, H. N. (Academic, New York), pp. 21–132.
22. Kumar, S., Tamura, K. & Nei, M. (1993) MEGA: *Molecular Evolutionary Genetics Analysis, Version 1* (Pennsylvania State Univ., University Park, PA).
23. Le Minor, L., Véron, M. & Popoff, M. (1982) *Ann. Microbiol. (Paris)* **133B**, 223–243.
24. Le Minor, L., Popoff, M. Y., Laurent, B. & Hermant, D. (1986) *Ann. Inst. Pasteur/Microbiol.* **137B**, 211–217.
25. Smith, N. H., Beltran, P. & Selander, R. K. (1990) *J. Bacteriol.* **172**, 2209–2216.
26. Li, J., Smith, N. H., Nelson, K., Crichton, P. B., Old, D. C., Whittam, T. S. & Selander, R. K. (1993) *J. Med. Microbiol.* **38**, 129–139.
27. Stephens, J. C. (1985) *Mol. Biol. Evol.* **2**, 539–556.
28. Ochman, H., Whittam, T. S., Caugant, D. A. & Selander, R. K. (1983) *J. Gen. Microbiol.* **129**, 2715–2726.
29. Ewing, W. H. (1986) *Edwards and Ewing's Identification of Enterobacteriaceae* (Elsevier, New York), 4th Ed.
30. Watanabe, H. & Okamura, N. (1991) in *The Prokaryotes*, eds. Ballow, A., Trüper, H. G., Dworkin, M., Harder, W. & Schleifer, K.-H. (Springer, New York), 2nd Ed., Vol. 3, pp. 2754–2759.
31. Karaolis, D. K. R., Lan, R. & Reeves, P. R. (1994) *J. Clin. Microbiol.* **32**, 796–802.
32. Stevenson, G., Neal, B., Liu, D., Hobbs, M., Packer, N. H., Batley, M., Redmond, J. W., Lindquist, L. & Reeves, P. (1994) *J. Bacteriol.* **176**, 4144–4156.
33. Brenner, D. J., Grimont, P. A. D., Steigerwalt, A. G., Fanning, G. R., Ageron, E. & Riddle, C. F. (1993) *Int. J. Syst. Bacteriol.* **43**, 645–658.
34. Lawrence, J. G., Ochman, H. & Hartl, D. L. (1991) *J. Gen. Microbiol.* **137**, 1911–1921.
35. Murray, N. E., Daniel, A. S., Cowan, G. M. & Sharp, P. M. (1993) *Mol. Microbiol.* **9**, 133–143.
36. Reeves, P. (1993) *Trends Genet.* **9**, 17–22.
37. Batchelor, R. A., Alifano, P., Biffali, E., Hull, S. I. & Hull, R. A. (1992) *J. Bacteriol.* **174**, 5228–5236.
38. Bastin, D. A., Stevenson, G., Brown, P. K., Haase, A. & Reeves, P. R. (1993) *Mol. Microbiol.* **7**, 725–734.
39. Keenleyside, W. J., Perry, M., Maclean, L., Poppe, C. & Whitfield, C. (1994) *Mol. Microbiol.* **11**, 437–448.
40. 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.
41. Whittam, T. S., Wolfe, M. L., Wachsmuth, I. K., Ørskov, F., Ørskov, I. & Wilson, R. A. (1993) *Infect. Immun.* **61**, 1619–1629.
42. Ørskov, F. & Ørskov, I. (1984) *Methods Microbiol.* **14**, 43–112.
43. Sugiyama, T., Kido, N., Komatsu, T., Ohta, M., Jann, K., Jann, B., Saeki, A. & Kato, N. (1994) *Microbiology* **140**, 59–71.
44. Jayaratne, P., Bronner, D., MacLachlan, P. R., Dodgson, C., Kido, N. & Whitfield, C. (1994) *J. Bacteriol.* **176**, 3126–3139.
45. Ochman, H. & Lawrence, J. G. (1994) 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), 2nd Ed., in press.
46. Sharp, P. M., Kelleher, J. E., Daniel, A. S., Cowan, G. M. & Murray, N. E. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 9836–9840.
47. Fitts, R. (1985) *Food Technology*, March 1985, pp. 95–102.
48. Groisman, E. A., Sturmoski, M. A., Solomon, F. R., Lin, R. & Ochman, H. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 1033–1037.
49. Van Vliet, F., Boyen, A. & Glansdorff, N. (1988) *Ann. Inst. Pasteur/Microbiol.* **139**, 493–496.
50. Hashimoto, Y., Li, N., Yokoyama, H. & Ezaki, T. (1993) *J. Bacteriol.* **175**, 4456–4465.
51. Waxin, H., Virlogeaux, L., Kolyva, S. & Popoff, M. Y. (1993) *Res. Microbiol.* **144**, 363–371.
52. Selander, R. K., Smith, N. H., Li, J., Beltran, P., Ferris, K. E., Kopecko, D. J. & Rubin, F. A. (1992) *J. Bacteriol.* **174**, 3587–3592.
53. Whittam, T. S., Ochman, H. & Selander, R. K. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1751–1755.
54. Maynard Smith, J., Smith, N. H., O'Rourke, M. & Spratt, B. G. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 4384–4388.
55. Achtman, M. & Pluschke, G. (1986) *Annu. Rev. Microbiol.* **40**, 185–210.
56. Musser, J. M., Bemis, D. A., Ishikawa, H. & Selander, R. K. (1987) *J. Bacteriol.* **169**, 2793–2803.
57. 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.
58. Achtman, M. (1994) *Mol. Microbiol.* **11**, 15–22.
59. Li, J., Nelson, K., McWhorter, A. C., Whittam, T. S. & Selander, R. K. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 2552–2556.
60. Caugant, D. A., Levin, B. R. & Selander, R. K. (1981) *Genetics* **98**, 467–490.