Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 6667–6671, August 1991
Genetics

Nucleotide polymorphism and evolution in the glyceraldehyde-3-phosphate dehydrogenase gene (*gapA*) in natural populations of *Salmonella* and *Escherichia coli*

(PCR/substitution rates/codon bias/enzyme electrophoresis/gene phylogeny)

KIMBERLYN NELSON, THOMAS S. WHITTAM, AND ROBERT K. SELANDER

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

Contributed by Robert K. Selander, May 9, 1991

ABSTRACT **Nucleotide sequences of the *gapA* gene, encoding the glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase, were determined for 16 strains of *Salmonella* and 13 strains of *Escherichia coli* recovered from natural populations. Pairs of sequences from strains representing the eight serovar groups of *Salmonella* differed, on average, at 3.8% of nucleotide sites and 1.1% of inferred amino acids, and comparable values for *E. coli* were an order of magnitude smaller (0.2% and 0.1%, respectively). The rate of substitution at synonymous sites was significantly higher for codons specifying the catalytic domain of the enzyme than for those encoding the NAD^+-binding domain, but the nonsynonymous substitution rate showed the opposite relationship. For *Salmonella*, statistical tests for nonrandom clustering of polymorphic sites failed to provide evidence that intragenic recombination or gene conversion has contributed to the generation of allelic diversity. The topology of a tree constructed from the *gapA* sequences was generally similar to that of phylogenetic trees of the strains based on multilocus enzyme electrophoresis, but the level of divergence of *gapA* in *Salmonella* group V from other *Salmonella* and *E. coli* strains is much greater than that indicated by DNA hybridization for the genome as a whole.**

Population genetic studies employing multilocus enzyme electrophoresis (1) to index allele diversity in chromosomal structural genes and measure linkage disequilibrium have indicated that natural populations of bacteria are highly polymorphic and have a clonal structure, which implies that the rate of homologous recombination is no greater than the mutation rate (2–4). At the same time, analyses of nucleotide sequence variation in individual genes or gene regions among isolates of several species of bacteria, including *Escherichia coli* (5–10), *Salmonella* (11, 12), and *Haemophilus influenzae* (13), have revealed mosaic patterns implicating intragenic recombination of horizontally transferred segments as a factor in the origin of new alleles.

To further study allelic diversity and other aspects of bacterial evolutionary genetics, we have generated a data base of nucleotide sequences of multiple genes in 29 diverse strains of *Salmonella* and *E. coli* selected to represent the ranges of genomic diversity in populations of these bacteria previously revealed by DNA hybridization experiments and multilocus enzyme electrophoresis. We here present the results of a comparative analysis of variation in the sequence of the gene (*gapA*) that encodes glyceraldehyde-3-phosphate dehydrogenase (GAPDH; EC 1.2.1.12), an enzyme of the glycolytic pathway that catalyzes the oxidation and phosphorylation of glyceraldehyde 3-phosphate to 1,3-bisphosphoglycerate (14).* We selected this enzyme for study because it is highly conserved in both bacteria and eukaryotes, its functional domains have been identified, its crystallin structure has been determined, it shows an unusually high degree of codon bias, and it is polymorphic for electromorph alleles, two of which are shared by *Salmonella* and *E. coli*. Our analysis has revealed that diversity at the *gapA* locus is an order of magnitude greater in *Salmonella* than in *E. coli*, that rates of substitution in the regions of the gene encoding the cofactor-binding and catalytic domains of the enzyme are not equivalent, and that, at least in *Salmonella*, allelic variation has been generated largely, if not exclusively, by point mutation rather than intragenic recombination.

MATERIALS AND METHODS

Bacterial Isolates. The *gapA* gene was sequenced in 16 strains of *Salmonella* and 13 strains of *E. coli*. From the *Salmonella* Reference Collection C (R.K.S., unpublished data), we selected 2 strains from each of the eight groups of serovars (I, II, IIIa, IIIb, IV, V, VI, and VII) that are currently distinguished by DNA hybridization, biotyping, and multilocus enzyme electrophoresis (15–18). The 16 strains, which have been analyzed electrophoretically for variation at 24 enzyme loci, are as follows (with the group indicated in parentheses): S3333 (representing the serovar *Salmonella typhi*) and S4194 [representing electrophoretic type TM 1, the predominant clone of serovar *Salmonella typhimurium* (19), of which laboratory strain LT2 is a member] (I), S2985 and S2993 (II), S2980 and S2983 (IIIa), S2978 and S2979 (IIIb), S3015 and S3027 (IV), S3041 and S3044 (V), S2995 and S3057 (VI), and S3013 and S3014 (VII).

From the *Escherichia coli* Reference Collection (20), we selected eight strains representing the four major phylogenetic lineages that have been identified by multilocus enzyme electrophoresis (38 loci) (3, 21, 22), as follows: EC10 and EC14 (representing group A, the lineage to which laboratory strain K-12 belongs); EC32, EC58, and EC70 (B1); EC52 and EC64 (B2); and EC40 (D). Five additional strains of *E. coli*, E3406, A8190, E2666-74, E830587, and E851819, selected from the research reference collection of T.S.W. to represent various GAPDH electromorphs, were also included in the analysis.

PCR Amplification and Nucleotide Sequencing. For each strain, we extracted genomic DNA (23) and amplified the coding region of the *gapA* gene (24) by PCR (25) with oligonucleotide primers designed from the published sequence for *E. coli* K-12 (26), as follows: 5′ primer (5′-TATGACTATCAAAGTAGGTAT-3′) and 3′ primer (5′-GTTGGAGTAACCGGTTTCGT-3′). The PCR product was treated with λ exonuclease to produce single-stranded DNA

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: GAPDH, glyceraldehyde-3-phosphate dehydrogenase.

*The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M66853 and M66882).

(27), and the resulting 924-base-pair fragment (representing 93% of the total coding region) was sequenced by the dideoxynucleotide chain-termination method, as supplied with Sequenase (United States Biochemical). Both orientations of the gene were sequenced by the use of six additional internal oligonucleotides, and the data were assembled and edited with SEQMAN and SEQMANED programs (DNASTAR, Madison, WI).

RESULTS

Nucleotide Polymorphism. The nucleotide sequence of a 924-base-pair segment of the *gapA* gene extending from codon 5 to codon 313 in the 330-amino acid coding region (26) was determined for 16 strains of *Salmonella* and 13 strains of *E. coli*. Of the total of 29 sequences obtained, 23 were distinctive; the sequences of the two strains of *Salmonella* group IIIb (S2978 and S2979) were identical, as were also those of the two strains of *Salmonella* group IV (S3015 and S3027). Of the *E. coli* sequences, EC10 and EC14 were identical to K-12 (26); and EC32, E3406, E851819, and A8190 were all identical.

Pairs of the 16 *Salmonella* sequences from strains representing eight serovar groups differed, on average, at 34.9 (3.8%) of the nucleotide sites and 3.5 (1.1%) of the inferred amino acid positions. Almost all of this variation is attributable to intergroup differences; within groups, the average pairwise difference was only 2.25 (0.24%) nucleotides and 0.125 (0.04%) amino acids. The level of variation among the 13 *E. coli* strains was an order of magnitude lower than in *Salmonella*, with an average pairwise difference of 2.4 (0.2%) nucleotides and 0.3 (0.1%) amino acids.

Alleles of the *Salmonella* group V strains (S3041 and S3044) were the most divergent, differing at 7.7–9.1% of nucleotide sites from the alleles of strains of other *Salmonella* groups and *E. coli*. In comparisons of group I strain S4194 (*S. typhimurium*), group V strain S3044, and strain EC10 of group A *E. coli* (Fig. 1), we identified 98 polymorphic nucleotide sites (occurring in 87 codons), which include 73% of the 134 polymorphic sites found in all 29 strains. There was a total of 118 polymorphic sites in the *Salmonella* genes, but only 12 polymorphic sites were detected among the *E. coli* sequences (data not shown). Four sites were polymorphic in both *Salmonella* and *E. coli*, and 8 sites were fixed for alternative bases in the two groups of bacteria.

Synonymous and Nonsynonymous Substitutions. GAPDH has two functional parts, a NAD^+-binding domain (amino acids 5–148) and a catalytic domain (amino acids 149–313) (26). We estimated the numbers of synonymous substitutions per 100 synonymous sites (d_S) and nonsynonymous substitutions per 100 nonsynonymous sites (d_N) (28, 29) separately for each domain and for both domains together for the *E. coli* sequences, the *Salmonella* sequences (with and without inclusion of group V), and all 29 sequences (Table 1). Overall there is clear evidence of selective constraint against amino acid replacement, because d_N is less than 8% of d_S in all comparisons except that for the binding domain of *E. coli*. There are, however, significant differences in both d_S and d_N between the domains. On average, d_S for sites in the catalytic domain is nearly two times greater than for sites in the binding domain, but d_N is greater in the binding domain than in the catalytic domain for comparisons involving all 29 sequences and the 14 *Salmonella* sequences (group V excluded). Exceptionally, the group V *Salmonella* strains showed eight amino acid replacements in the catalytic domain but only four replacements in the binding domain. For *E. coli*, values of d_N, which were very small, were nearly equal in the two regions of the gene.

Distribution of Polymorphic Sites. The methods of Sawyer (30) and Stephens (31) were used to test for nonrandom clustering of polymorphic synonymous sites, a pattern that may be indicative of intragenic recombination or gene conversion. For the 96 polymorphic synonymous sites in the 14 distinctive *Salmonella* alleles, neither method detected nonrandom clustering. In Sawyer's test, for example, the sum of the squared condensed fragment lengths (SSCF) was 77,446, and the maximum length of condensed fragments (MCF) was 66; the resulting probabilities (P) of obtaining values equal to or greater than the observed values in comparison to the null distribution generated by computer sampling were 0.41 and 0.97, respectively. In a comparable test in which only one allele from each of the eight serovar groups was considered, SSCF = 6622 (P = 0.20) and MCF = 30 (P = 0.40). For *E. coli*, with a total of only 12 polymorphic sites, meaningful testing of the distribution pattern was not feasible.

Correlation of Electromorphs and Nucleotide Sequences. Five distinctive electromorphs of GAPDH have been distinguished among the 29 strains of *Salmonella* and *E. coli* (K.N., unpublished data). In *Salmonella*, electromorph A occurs in all strains of groups I, II, IIIa, IIIb, and IV, and each of three

Position	37	38	56	69	76	77	91	92	95	96	97	98	101	104	105	107	111	112	114	118	119	124	125	126	128	134	136	137	138
	Ala	Glu	Thr	Lys	Glu	Arg	Val	Val	Ala	Thr	Gly	Ile	Thr	Thr	Ala	Lys	Thr	Gly	Lys	Leu	Thr	Asn	Thr	Pro	Phe	Phe	Lys	Tyr	Glu
S4194	GCA	GAG	ACT	AAA	GAA	CGC	GTA	GTG	GCT	ACC	GGT	ATC	ACT	ACC	GCG	AAA	ACT	GGC	AAA	CTG	ACG	AAC	ACC	CCA	TTT	TTT	AAA	TAC	GAA
												Leu					Ala			Met							Thr		Ala
S3044	...	...	...	..G	..G	...	..C	..T	..G	..T	..C	C.G	..C	..T	..T	..G	G..	...	...	A..	..C	..T	...	..G	..C	..C	.CT	...	.C.
		Asp										Leu					Ala			Met									Ala
EC10	..T	..T	..C	...	...	..T	..T	..C	..A	..T	...	C.G	...	..T	..T	...	G..	..T	..G	A..	..T	...	..T	..G	..C	..C	...	..T	.CT

Position	145	147	151	155	158	160	163	166	168	170	173	174	177	178	180	181	186	187	196	199	200	208	209	210	211	213	219	222	223
	Ser	Ala	Thr	Ala	Ala	Val	Asp	Gly	Ile	Gly	Thr	Thr	Ala	Thr	Ala	Thr	Asp	Gly	Gly	Ala	Ser	Thr	Gly	Ala	Ala	Ala	Pro	Asn	Gly
S4194	TCC	GCT	ACC	GCG	GCT	GTT	GAC	GGC	ATC	GGT	ACT	ACT	GCG	ACT	GCA	ACC	GAC	GGC	GGC	GCA	TCT	ACT	GGC	GCT	GCT	GCG	CCG	AAT	GGC
									Val												Ala								
S3044	..T	..A	..T	...	..A	..G	...	..T	G.T	..C	..C	...	..C	..C	..T	..T	..T	..T	..T	...	G..	..G	..T	..A	..G	..A	...	..C	..T
EC10	...	...	...	..T	...	...	..T	...	...	...	..C	..C	..T	...	..T	..T	..T	...	...	..T	..C	..C	..T	...	...	..T	..A	...	...

Position	231	234	238	239	243	245	249	250	251	252	253	254	255	256	257	258	259	263	266	273	276	278	279	287	292	296	297	299	308
	Arg	Thr	Ser	Val	Thr	Arg	Ala	Ala	Thr	Tyr	Glu	Gln	Ile	Lys	Ala	Ala	Val	Ala	Glu	Gly	Asp	Val	Val	Val	Phe	Ala	Gly	Ala	Val
S4194	CGC	ACT	TCC	GTT	ACC	CGT	GCG	GCT	ACC	TAC	GAG	CAG	ATC	AAA	GCT	GCT	GTT	GCC	GAA	GGT	GAC	GTT	GTA	GTA	TTC	GCA	GGC	GCG	GTC
									Ser			Glu			Lys		Met	Ser	Ala										
S3044	..T	..G	...	..C	..T	..C	...	...	T.T	..T	..A	G..	..T	..G	AAA	..A	A.G	T..	.C.	..C	..T	..A	..T	...	..T	...	..T	..A	...
EC10	...	..C	..T	..A	...	...	..T	..A	..T	...	...	...	...	...	...	..C	...	..T	...	..C	..T	..A	...	..T	...	..T	..T	..T	..A

FIG. 1. Distribution of polymorphic sites in *gapA* for strains S4194 (*S. typhimurium*, representing group I *Salmonella*), S3044 (group V *Salmonella*), and EC10 (which is identical in sequence to laboratory strain K-12). Among these three strains, there are 87 polymorphic codons whose positions in the inferred polypeptide of 330 amino acids are numbered. Represented among these three sequences are 73% of the 134 polymorphic sites detected among the 29 sequences analyzed.

Table 1. Nucleotide substitutions in regions of *gapA* encoding two functional domains of GAPDH

Species	Nucleotide substitutions, no. per 100 sites					
	NAD^+-binding domain*		Catalytic domain*		Both domains	
	d_S	d_N	d_S	d_N	d_S	d_N
E. coli (13)	0.46 ± 0.26	0.09 ± 0.07	1.01 ± 0.47	0.08 ± 0.05	0.78 ± 0.28	0.09 ± 0.04
Salmonella (14)†	5.33 ± 1.23	0.26 ± 0.16	14.68 ± 2.10	0.00 ± 0.02	10.31 ± 1.24	0.12 ± 0.08
Salmonella (16)	8.91 ± 1.64	0.58 ± 0.22	21.60 ± 2.48	0.64 ± 0.21	15.55 ± 1.49	0.61 ± 0.15
E. coli plus *Salmonella* (29)	13.57 ± 2.51	0.98 ± 0.37	22.35 ± 2.68	0.40 ± 0.12	18.22 ± 1.85	0.67 ± 0.18

Numbers in parentheses are numbers of sequences analyzed.
*Numbers of substitutions are based on 429 nucleotides of codons 5–148 in the NAD^+-binding domain and 495 nucleotides of codons 149–313 in the catalytic domain.
†Group V strains were excluded.

other electromorphs is confined to a single group (C to V, D to VI, and F to VII). D is the electromorph of all strains of *E. coli* except E830587 (B) and E2666-74 (F).

Among the inferred amino acid sequences of all strains, there are 16 substitutions, of which 7 produce charge changes (Table 2). Every electromorphic difference involves at least one charge change, and no charge change was observed that does not result in a shift in the electrophoretic mobility of the protein. *Salmonella* and *E. coli* share two electromorphs, D and F, and in each case the similarity in net charge depends on substitutions at different codon positions, 85 and 138 for D and 85 and 266 for F. Electromorph C of the group V *Salmonella* differs from the common electromorph A by 5 amino acid charge changes, which, however, yield a net loss of only a single negative charge (Table 2); these alleles also differ by 7 amino acid replacements that do not affect the net charge (Fig. 1). Among strains of the same species, the same electromorph may vary markedly at synonymous nucleotide sites; for example, the A electromorphs of *Salmonella* groups I and IIIa differ by 33 synonymous substitutions.

Evolutionary Tree for *gapA* Sequences. An evolutionary tree constructed by the neighbor-joining method (32) on the basis of polymorphic nucleotide sites is shown in Fig. 2, together with comparable trees summarizing the genomic relationships of the 16 *Salmonella* strains and the 8 strains from the *Escherichia coli* Research Collection, as indexed by multilocus enzyme electrophoresis. The *gapA* tree was rooted by the use of sequences of the *gapA* gene of a strain of *Klebsiella pneumoniae* (K.N., unpublished data) and the *gap* gene of *Bacillus subtilis* (33) as outgroups; and the significance of the branching order was determined by bootstrap analysis of 1000 computer-generated trees.

At the first node in the *gapA* tree, group V diverges from *E. coli* and the other groups of *Salmonella*. The topological connection of the enzyme electrophoresis trees for *Salmonella* and *E. coli* has not been determined.

For *Salmonella*, there is only one notable difference in branching order between the two trees: group VII clusters with group IIIB in the enzyme electrophoresis tree but with group IV in the *gapA* tree; however, the latter association was represented in only 53% of bootstrap trees. The *gapA* alleles of *E. coli* were clearly derived from a single common ancestor; and although the number of substitutions among the strains is too small to permit reliable resolution of sublineages, it is noteworthy that there is close similarity in the topologies of the *gapA* and enzyme electrophoresis trees.

A notable feature of the trees is that many pairs or groups of strains are identical or nearly so in *gapA* sequence but divergent in overall genomic character. The most striking case is that of the two *Salmonella* group I strains (S4194 and S3333), which differ at only two nucleotide sites in *gapA* sequence but are separated by a genetic distance of 0.61 on the basis of multilocus enzyme electrophoresis.

Table 2. Distribution of amino acid charge changes in *gapA* sequences of *Salmonella* and *E. coli*

Electromorph	Amino acid at position indicated						
	85	136	138	254	257	265	266
Salmonella							
A	Asp^-	Lys^+	Glu^-	Gln	Ala	Gly	Glu^-
C	Asp^-	Thr	Ala	Glu^-	Lys^+	Gly	Ala
D	Asn	Lys^+	Glu^-	Gln	Ala	Gly	Glu^-
F	Asn	Lys^+	Ala	Gln	Ala	Gly	Glu^-
E. coli							
B	Asp^-	Lys^+	Ala	Gln	Ala	Asp^-	Glu^-
D	Asp^-	Lys^+	Ala	Gln	Ala	Gly	Glu^-
F	Asp^-	Lys^+	Ala	Gln	Ala	Gly	Ala

DISCUSSION

Evolution of GAPDH. As a key enzyme in the glycolytic pathway, GAPDH is expected to be under strong selective constraint for amino acid replacement, and this clearly is the case for the molecule as a whole, inasmuch as nonsynonymous changes in the gene occur with a frequency of less than 8% of that of synonymous substitutions. The nucleotide and amino acid sequence differences for *gapA* between *E. coli* EC10 and *S. typhimurium* S4194 are 6.0% and 1.3%, respectively; these values are similar to those reported for the B subunit of RNA polymerase (6.1% and 1.4%, respectively) of laboratory strains *E. coli* K-12 and *S. typhimurium* LT2 and are well below the averages for 55 genes in these strains (14% and 7.7%, respectively).

For *E. coli*, Branlant and Branlant (26) have shown that the two functional domains of GAPDH exhibit different degrees of amino acid sequence similarity with GAPDH in eukaryotes and bacteria. Whereas the NAD^+-binding domain of *E. coli* is 45–60% similar to comparable sequences of both eukaryotes and other bacteria, the catalytic domain is 74–80% similar to eukaryotic domains but only 50–60% similar to other bacterial sequences. In our analysis, this difference between the domains is reflected in the relative rates of amino acid replacement, with nonsynonymous substitutions, although infrequent, occurring, in general, more often in the binding domain than in the catalytic domain. However, there is an exception: strains of group V *Salmonella* have more amino acid replacements in the catalytic domain than in the binding domain.

Synonymous substitutions show a different pattern, with the frequency in the catalytic domain being roughly twice that in the binding domain. However, the estimated numbers of synonymous changes (d_S) shown in Table 1 were obtained by the method of Nei and Gojobori (28), which does not take into account nonrandom patterns of synonymous codon usage and gives equal weight to all possible pathways between codons with multiple nucleotide differences. Consequently, this method may yield poor estimates of the synonymous

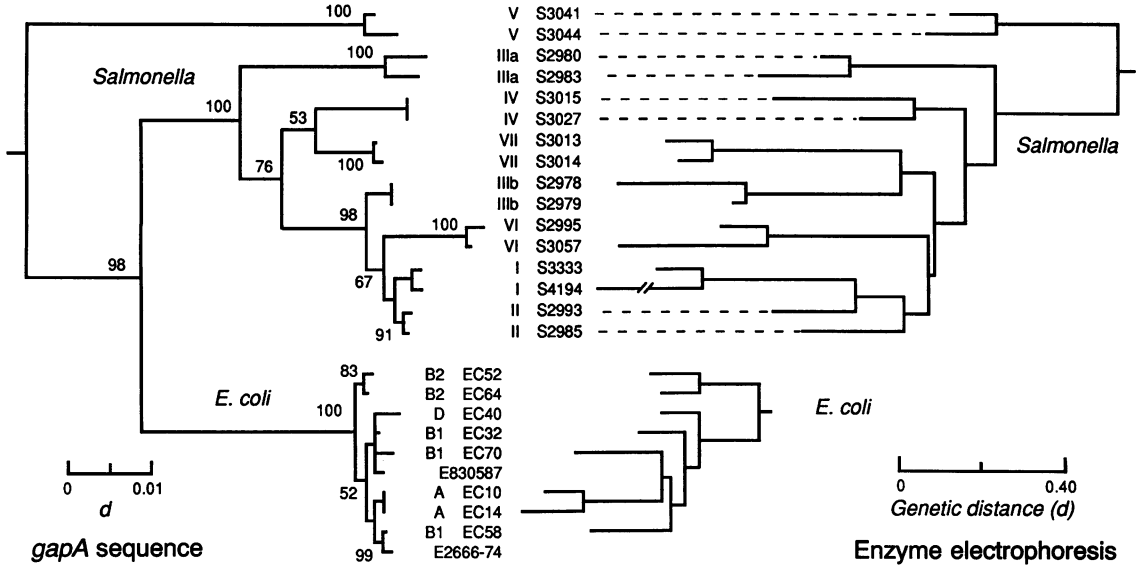

FIG. 2. (*Left*) Phylogenetic tree of *gapA* genes of 16 strains of *Salmonella* and 13 strains of *E. coli*, constructed by the neighbor-joining method (32) from a matrix of pairwise differences per site for 134 polymorphic nucleotide sites. The number adjacent to a node indicates the percentage of bootstrap trees that contained that node. The sequences of *E. coli* strains E3406, E851819, and A8190 (not shown) were identical to that of EC32. (*Right*) Phylogenetic trees for 16 strains of *Salmonella* and 8 strains of *E. coli* from the *Escherichia coli* Research Collection, based on electrophoretically demonstrable variation in 24 and 38 enzymes, respectively. The neighbor-joining trees were constructed from a matrix of pairwise values of Nei's genetic distance (*d*). The branch for strain S4194 was truncated to half its actual length.

substitution rate for genes that exhibit strong codon bias; and the *gapA* gene is highly codon biased, with an average codon adaptation index (34) of 0.81 (range, 0.77–0.83). To assess the influence of codon bias on estimates of the rate of synonymous substitution, we used Lewontin's method (35) to calculate synonymous changes for 2-fold and 4-fold redundant sites separately. As shown in Fig. 3, these estimates are highly correlated with d_S for all pairwise comparisons among the 23 distinctive *gapA* alleles: values of d_S are slightly smaller than Lewontin's measure for 4-fold sites and larger than his measure for 2-fold sites. Between the two domains, there is no difference in the proportions of substitutions at 2-fold and 4-fold redundant sites. Of all substitutions in the binding domain, 32% occurred in 2-fold sites and 63% were in 4-fold sites; and in the catalytic domain, the comparable proportions were 20% and 68%, respectively. These results

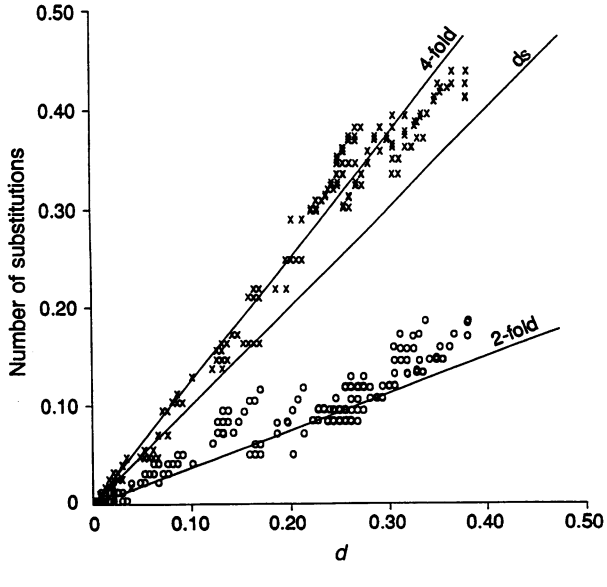

FIG. 3. Correlation between d_S and Lewontin's estimates (35) of the synonymous rate of substitution for 2-fold and 4-fold redundant sites. *d*, Synonymous changes per synonymous site.

indicate that d_S is a reliable measure of the relative degree of genetic divergence at synonymous sites for genes with high codon bias when $d_S < 0.5$.

Evolutionary Relationships. Doolittle *et al.* (36) postulated that an ancestor of *E. coli* acquired the *gapA* gene from a eukaryote. Comparative analysis of our data and *gap* sequences of other bacteria and various eukaryotes clearly indicates that the *gapA* genes of *Salmonella*, *E. coli*, and the outgroup *K. pneumoniae* diverged from a common ancestral gene, irrespective of its origin.

Both DNA hybridization (15, 16) and multilocus enzyme electrophoresis (refs. 17 and 18; see Fig. 2) have demonstrated that group V is the most distinctive of the eight serovar groups of *Salmonella*. In DNA hybridization experiments, values of relative reassociation at 75°C reportedly are in the range of 58–69% for comparisons of strains of groups I, II, IIIa, IIIb, IV, and VI but in the range of 41–49% for comparisons of strains of group V with those of other groups (15, 16).

Considered alone, the topology of the *gapA* tree (Fig. 2) might be interpreted as evidence that group V is phylogenetically more distant from other groups of *Salmonella* than the latter are from *E. coli*. But single gene trees cannot be equated with phylogenetic trees of genomes, and data from DNA hybridization experiments and the comparative sequencing of other genes indicate that strains of group V are in their overall genetic character not as divergent as their *gapA* sequences would suggest. Reported values of relative reassociation of DNA from *E. coli* K-12 with that of strains of the *Salmonella* groups, including group V, are in the range of only 11–26% at 75°C (15, 37). In *gapA* sequence, group V is nearly as different from other *Salmonella* as is *Klebsiella*, which is not closely related to either *Salmonella* or *E. coli* (7% and 18% relative binding of *K. pneumoniae* DNA at 75°C with that of *E. coli* K-12 and *S. typhimurium* LT2, respectively) (ref. 38; see also ref. 39). Results of comparative sequencing of the proline permease gene (*putP*) are more in line with the relationships of group V to other *Salmonella* and *E. coli* indicated by DNA hybridization. Thus, for example, group V differs from group I at 5% of nucleotide sites, whereas both group I and V differ from *E. coli* at 16% of sites

(K.N., unpublished data). We conclude that the group V lineage either experienced an unusually rapid rate of evolution of *gapA* or acquired part or all of the gene by horizontal transfer from an as-yet-unidentified source outside of *Salmonella*.

Divergence Times. The average rate of nonsynonymous substitution per site in the *gap* gene of eukaryotes has been estimated as 0.14×10^{-9} per year on the basis of calibration with several points in the fossil record (40). In application of this rate to our *gapA* sequences, the time since divergence of the group I *Salmonella* (*S. typhimurium* and *S. typhi*) from *E. coli* is estimated to be 31 ± 12.5 million years (Myr). All things considered, this value is remarkably similar to estimates of 37 ± 26 Myr and 30 ± 10 Myr for the time of divergence of these bacteria derived from studies of rates of sequence evolution in the 5S RNA gene (41) and the glutamine synthetase gene (42), respectively. All these dates are much later than the 120–160 Myr proposed by Ochman and Wilson (43), based in part on the supposition that *E. coli* evolved from a common ancestor with *Salmonella* at the time of origin of the mammals.

We thank J. M. Gearhart, M. Hughes, and S. A. Plock for technical assistance and S. W. Schaeffer for helpful comments on the manuscript. This research was supported by Grants AI-22144 (to R.K.S.) and AI-24566 and AI-00964 (to T.S.W.) from the National Institutes of Health.

1. Selander, R. K., Caugant, D. A., Ochman, H., Musser, J. M., Gilmour, M. N. & Whittam, T. S. (1986) *Appl. Environ. Microbiol.* **51,** 873–884.
2. Whittam, T. S., Ochman, H. & Selander, R. K. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1751–1755.
3. 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), Vol. 2, pp. 1625–1648.
4. 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.
5. Milkman, R. & Crawford, I. R. (1983) *Science* **221,** 378–380.
6. Sawyer, S., Dykhuizen, D. & Hartl, D. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 6225–6228.
7. Barcak, G. J. & Wolf, R. E., Jr. (1988) *J. Bacteriol.* **170,** 372–379.
8. DuBose, R. F., Dykhuizen, D. E. & Hartl, D. L. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 7036–7040.
9. Stoltzfus, A., Leslie, J. F. & Milkman, R. (1988) *Genetics* **120,** 345–358.
10. Milkman, R. & Bridges, M. M. (1990) *Genetics* **126,** 505–517.
11. Smith, N. H., Beltran, P. & Selander, R. K. (1990) *J. Bacteriol.* **172,** 2209–2216.
12. Smith, N. H. & Selander, R. K. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 956–960.
13. Kroll, J. S. & Moxon, E. R. (1990) *J. Bacteriol.* **172,** 1374–1379.
14. Harris, J. I. & Waters, M. (1976) in *The Enzymes*, ed. Boyer, P. D. (Academic, New York), Vol. 13, pp. 1–49.
15. Le Minor, L., Veron, M. & Popoff, M. (1982) *Ann. Microbiol. (Paris)* **133B,** 223–243.
16. Le Minor, L., Popoff, M. Y., Laurent, B. & Hermant, D. (1986) *Ann. Microbiol. (Paris)* **137B,** 211–217.
17. Reeves, M. W., Evins, G. M., Heiba, A. A., Plikaytas, B. D. & Farmer, J. J., III (1989) *J. Clin. Microbiol.* **27,** 313–320.
18. 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.
19. 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.
20. Ochman, H. & Selander, R. K. (1984) *J. Bacteriol.* **157,** 690–693.
21. Goullet, P. H. & Picard, B. (1989) *J. Gen. Microbiol.* **135,** 135–142.
22. Herzer, P. J., Inouye, S., Inouye, M. & Whittam, T. S. (1990) *J. Bacteriol.* **171,** 6175–6181.
23. Wilson, K. (1990) in *Current Protocols in Molecular Biology*, eds. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (Wiley, New York), Vol. 1, pp. 2.4.1–2.4.2.
24. Alefounder, P. R. & Perham, R. N. (1989) *Mol. Microbiol.* **3,** 723–732.
25. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239,** 487–491.
26. Branlant, G. & Branlant, C. (1985) *Eur. J. Biochem.* **150,** 61–66.
27. Higuchi, R. G. & Ochman, H. (1989) *Nucleic Acids Res.* **17,** 5865.
28. Nei, M. & Gojobori, T. (1986) *Mol. Biol. Evol.* **3,** 418–426.
29. Nei, M. & Jin, L. (1989) *Mol. Biol. Evol.* **6,** 290–300.
30. Sawyer, S. (1989) *Mol. Biol. Evol.* **6,** 526–538.
31. Stephens, J. C. (1985) *Mol. Biol. Evol.* **2,** 539–556.
32. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4,** 406–425.
33. Viaene, A. & Dhaese, P. (1989) *Nucleic Acids Res.* **17,** 1251.
34. Sharp, P. M. & Li, W.-H. (1987) *Nucleic Acids Res.* **15,** 1281–1295.
35. Lewontin, R. C. (1989) *Mol. Biol. Evol.* **6,** 15–32.
36. Doolittle, R. F., Feng, D. F., Anderson, K. L. & Alberro, M. R. (1990) *J. Mol. Evol.* **31,** 383–388.
37. Crosa, J. H., Brenner, D. J., Ewing, W. H. & Falkow, S. (1973) *J. Bacteriol.* **115,** 307–315.
38. Brenner, D. J. (1973) *Int. J. Syst. Bacteriol.* **23,** 298–307.
39. Riley, M. & Anilionis, A. (1980) *J. Bacteriol.* **143,** 366–376.
40. Martin, W., Gierl, A. & Saedler, H. (1989) *Nature (London)* **339,** 46–48.
41. Hori, H. & Osawa, S. (1978) *J. Bacteriol.* **133,** 1089–1095.
42. Pesole, G., Bossetti, M. P., Lanave, C., Preparata, G. & Saccone, C. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 522–526.
43. Ochman, H. & Wilson, A. C. (1987) *J. Mol. Evol.* **26,** 74–86.