Molecular and evolutionary relationships among enteric bacteria

JEFFREY G. LAWRENCE, HOWARD OCHMAN and DANIEL L. HARTL*

Department of Genetics, Box 8232, Washington University School of Medicine, St Louis, MO 63110, USA

(Received 19 November 1990; revised 3 April 1991; accepted 10 April 1991)

Classification of bacterial species into genera has traditionally relied upon variation in phenotypic characteristics. However, these phenotypes often have a multifactorial genetic basis, making unambiguous taxonomic placement of new species difficult. By designing evolutionarily conserved oligonucleotide primers, it is possible to amplify homologous regions of genes in diverse taxa using the polymerase chain reaction and determine their nucleotide sequences. We have constructed a phylogeny of some enteric bacteria, including five species classified as members of the genus *Escherichia*, based on nucleotide sequence variation at the loci encoding glyceraldehyde-3-phosphate dehydrogenase and outer membrane protein 3A, and compared this genealogy with the relationships inferred by biotyping. The DNA sequences of these genes defined congruent and robust phylogenetic trees indicating that they are an accurate reflection of the evolutionary history of the bacterial species. The five species of *Escherichia* were found to be distantly related and, contrary to their placement in the same genus, do not form a monophyletic group. These data provide a framework which allows the relationships of additional species of enteric bacteria to be inferred. These procedures have general applicability for analysis of the classification, evolution, and epidemiology of bacterial taxa.

Introduction

Classification of bacteria into species, genera and higher categories has been a historically enigmatic issue. While an abundance of phenotypic characters allows rigorous taxonomic treatment of many eukaryotic phyla, few readily available characters aid in the unambiguous classification of bacterial strains. Accurate classification of bacterial species is necessary for clinical and epidemiological studies, and a taxonomy which accurately reflects the evolutionary relationships among species provides a framework for elucidating the evolution of genes, gene families and metabolic pathways, for interpreting the distribution of extrachromosomal elements, and for inferring potential transfer of genetic material between species.

Traditionally, metabolic characteristics, such as the ability to metabolize certain carbon sources, or the production or catabolism of various biochemical compounds, have been used to classify bacterial strains into species and species groups on purely phenetic grounds. More recently, total DNA relatedness has been proposed as a more reliable method for grouping conspecific bacterial strains (e.g. Brenner, 1978). While either method can define species boundaries, criteria for grouping species into genera are less clear. Biochemical and metabolic phenotypes often have a multifactorial genetic basis, that is, multiple genes are responsible for their expression. Owing to convergent evolutionary events in distantly related lineages, these characters cannot be used reliably for a hierarchical grouping of species in a cladistic manner (one based on the identification of shared derived characteristics). DNA hybridization methods, while sufficient to define the range of variation expected for a bacterial species, do not provide an accurate assessment of relatedness allowing the grouping of congeneric species, and the placement of species into genera has relied upon characteristic metabolic features which are neither rigorously defined nor statistically testable (Brenner et al., 1978, 1980).

To eliminate such problems in the classification of bacterial species, their genetic relationships must be assessed beyond the accuracy allowed by total DNA relatedness. The DNA sequences of slowly evolving loci can provide the information necessary to unambiguously classify taxa above the species level and avoid the
inhomogeneities intrinsic to epistatic phenotypic traits. The necessary DNA sequences are readily obtainable by the polymerase chain reaction (PCR) (Saiki et al., 1985, 1988). Classification based on patterns of nucleotide substitutions allows statistical testing of the inferred evolutionary relationships, while statistical tests of phylogenies based on weighted phenotypic characters are either unavailable or unreliable. Escherichia coli is one of the most intensively studied organisms in terms of its genetics, physiology and molecular biology. Since 1973, an additional four species of Escherichia have been described: Escherichia blattae (Burgess et al., 1973), Escherichia fergusonii (Farmer et al., 1985), Escherichia hermannii (Brenner et al., 1982a) and Escherichia vulneris (Brenner et al., 1982b). With the exception of E. blattae, DNA relatedness was used to classify each group as a distinct species. The classification of all four species in this genus was based upon overall phenotypic similarity, and each exhibited a combination of metabolic characters atypical for this genus.

By designing oligonucleotide primers homologous to evolutionarily conserved sequences, it is possible to amplify homologous regions of ubiquitous genes in diverse taxa using PCR and to determine their nucleotide sequences (Kocher et al., 1989; Taylor & Kreitman, 1989). We have determined the DNA sequences of major portions of the gap and ompA genes, which encode glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and outer membrane protein 3A, respectively, of strains representative of eleven species of enteric bacteria, including the five species of Escherichia. The inferred genetic relationships were used to test the correlation of the phenotypic characters defining the genus Escherichia.

**Methods**

**Bacterial strains and genetic loci.** Chromosomal DNA was isolated from the eighteen strains of enteric bacteria listed in Table 1, by methods described previously (Sawyer et al., 1987). The gap locus encodes GAPDH and is located at 39-3 min on the Escherichia coli linkage map (Bachmann, 1990). The ompA locus encodes outer membrane protein 3A and is located at 21-8 min on the linkage map (Bachmann, 1990). Sequences homologous to the E. coli gap and ompA loci were verified in each of the strains listed in Table 1 by Southern blot genomic DNA hybridization analysis (data not shown). Using the published nucleotide sequences of the gap locus from E. coli, Bacillus stearothermophilus and eukaryotes (Branlant & Branlant, 1985; Branlant et al., 1989; Stone et al., 1985), and of the ompA locus from E. coli and Serratia marcescens (Beck & Bremer, 1980; Braun & Cole, 1984), oligonucleotide primers were synthesized complementary to highly conserved regions of these genes containing low codon redundancy. These primers allow the PCR amplification of a 906 bp fragment encompassing 91 % of the gap locus and a 711 bp fragment spanning 68 % of the ompA locus. The gap oligonucleotides were: dACTATCAAGTAGGATCATCAA, forward strand at nucleotide 4; dGGATTTGGCCGATCGGCCG, forward strand at nucleotide 25; dCGGAATCTCCTGTGTACAC, reverse strand at nucleotide 910. The ompA oligonucleotides were: dAAAGCCTCAGGCGCTTCA, forward strand at nucleotide 280; dGCCGCTGTACACAATCTT, reverse strand at nucleotide 1031.

**Template preparation.** PCR (Saiki et al., 1985, 1988) was performed in a buffer containing 50 mM-KCl, 10 mM-Tris (pH 8.4), 2.5 mM-MgCl2, 0.01 % gelatin, 800 µM-dNTP, 25 U Taq polymerase ml-1 (Cetus), 2 ng of each oligonucleotide primer µl-1 and 10-50 ng chromosomal DNA. The amplified sample was size-fractionated by agarose gel electrophoresis and DNA fragments of the appropriate size were isolated on Schleicher and Schuell NA45 DEAE paper. The paper-bound DNA was briefly washed in a buffer [150 mM-NaCl, 50 mM-Tris (pH 8.0), 10 mM-EDTA] at room temperature and eluted in a solution of 1 M-NaCl, 50 mM-Tris (pH 8.0), 10 mM-EDTA at 68°C for 30 min. The DNA was serially extracted with buffer-saturated phenol and chloroform/isomyl alcohol (24:1, v/v), and precipitated with 2.5 vols 100% ethanol in the presence of a linear polyacrylamide carrier (Galliard & Strauss, 1990). DNA was recovered by centrifugation, rinsed with 70% (v/v) ethanol to remove salts and dried in a vacuum desiccator prior to resuspension in 10 mM-Tris, 1 mM-EDTA.

**DNA sequencing.** Each DNA template (100 ng) was denatured in 0.2 M-NaOH for 5 min, neutralized by the addition of one-third vol. 3 M-potassium acetate (pH 4.6) and precipitated with 2.5 vols of 100% ethanol in the presence of additional polyacrylamide carrier. Following incubation at -20°C for 30 min, DNA was recovered by centrifugation as described above and sequenced with Sequenase 2.0 (United States Biochemicals) according to the manufacturer’s instructions. The sequences of both strands of each DNA template were determined using internal oligonucleotide primers (sequences determined by these methods yield over 99.9% accuracy). DNA sequence analysis was facilitated by the PAP (version 2.4, D. Swofford) and MacClade

---

**Table 1. Strains analysed for gap and ompA sequences**

<table>
<thead>
<tr>
<th>Strain</th>
<th>gap</th>
<th>CAI*</th>
<th>Mol% G+C</th>
<th>CAI*</th>
<th>Mol% G+C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citrobacter freundii OS60†</td>
<td>0.76</td>
<td>51.8</td>
<td>0.82</td>
<td>49.6</td>
<td></td>
</tr>
<tr>
<td>Enterobacter aerogenes E482‡</td>
<td>0.73</td>
<td>53.6</td>
<td>0.75</td>
<td>50.2</td>
<td></td>
</tr>
<tr>
<td>Escherichia blattae ATCC 29907†</td>
<td>0.81</td>
<td>54.5</td>
<td>0.80</td>
<td>52.4</td>
<td></td>
</tr>
<tr>
<td>E. blattae ATCC 33430</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Escherichia coli K12§</td>
<td>0.79</td>
<td>54.7</td>
<td>0.83</td>
<td>50.1</td>
<td></td>
</tr>
<tr>
<td>Escherichia fergusonii ATCC 35469†</td>
<td>0.82</td>
<td>53.5</td>
<td>0.83</td>
<td>49.4</td>
<td></td>
</tr>
<tr>
<td>E. fergusonii ATCC 35471</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Escherichia hermannii ATCC 33650(§)</td>
<td>0.81</td>
<td>53.1</td>
<td>0.83</td>
<td>50.5</td>
<td></td>
</tr>
<tr>
<td>E. hermannii ATCC 33652</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Escherichia vulneris ATCC 29943†</td>
<td>0.81</td>
<td>55.6</td>
<td>0.83</td>
<td>51.7</td>
<td></td>
</tr>
<tr>
<td>E. vulneris ATCC 33821</td>
<td>0.77</td>
<td>53.9</td>
<td>0.82</td>
<td>51.2</td>
<td></td>
</tr>
<tr>
<td>E. vulneris ATCC 33822</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Klebsiella pneumoniae LD119</td>
<td>0.82</td>
<td>56.5</td>
<td>0.85</td>
<td>52.6</td>
<td></td>
</tr>
<tr>
<td>Salmonella typhimurium LT2§</td>
<td>0.74</td>
<td>54.2</td>
<td>0.78</td>
<td>50.8</td>
<td></td>
</tr>
<tr>
<td>Serratia marcescens ATCC 13880†</td>
<td>0.73</td>
<td>53.5</td>
<td>0.69</td>
<td>53.3</td>
<td></td>
</tr>
<tr>
<td>Serratia odorifera ATCC 3307†</td>
<td>0.73</td>
<td>53.7</td>
<td>0.69</td>
<td>52.4</td>
<td></td>
</tr>
</tbody>
</table>

* Codon adaptation index (Sharp & Li, 1987b).
† Clinical isolate (Bergström et al., 1984).
§ Clinical isolate from Wiedermann, Bonn, Germany.
© Laboratory collection.
||Clinical isolate from S. Normark, Östersund, Sweden.
§ Type strain.

dCCGAACTCTCGTTGTGTACCA, reverse strand at nucleotide 910. The ompA oligonucleotides were: dAAAGCCTCAGGCGCTTCA, forward strand at nucleotide 280; dGCCGCTGTACACAATCTT, reverse strand at nucleotide 1031.

**Template preparation.** PCR (Saiki et al., 1985, 1988) was performed in a buffer containing 50 mM-KCl, 10 mM-Tris (pH 8.4), 2.5 mM-MgCl2, 0.01 % gelatin, 800 µM-dNTP, 25 U Taq polymerase ml-1 (Cetus), 2 ng of each oligonucleotide primer µl-1 and 10-50 ng chromosomal DNA. The amplified sample was size-fractionated by agarose gel electrophoresis and DNA fragments of the appropriate size were isolated on Schleicher and Schuell NA45 DEAE paper. The paper-bound DNA was briefly washed in a buffer [150 mM-NaCl, 50 mM-Tris (pH 8.0), 10 mM-EDTA] at room temperature and eluted in a solution of 1 M-NaCl, 50 mM-Tris (pH 8.0), 10 mM-EDTA at 68°C for 30 min. The DNA was serially extracted with buffer-saturated phenol and chloroform/isomyl alcohol (24:1, v/v), and precipitated with 2.5 vols 100% ethanol in the presence of a linear polyacrylamide carrier (Galliard & Strauss, 1990). DNA was recovered by centrifugation, rinsed with 70% (v/v) ethanol to remove salts and dried in a vacuum desiccator prior to resuspension in 10 mM-Tris, 1 mM-EDTA.

**DNA sequencing.** Each DNA template (100 ng) was denatured in 0.2 M-NaOH for 5 min, neutralized by the addition of one-third vol. 3 M-potassium acetate (pH 4.6) and precipitated with 2.5 vols of 100% ethanol in the presence of additional polyacrylamide carrier. Following incubation at -20°C for 30 min, DNA was recovered by centrifugation as described above and sequenced with Sequenase 2.0 (United States Biochemicals) according to the manufacturer’s instructions. The sequences of both strands of each DNA template were determined using internal oligonucleotide primers (sequences determined by these methods yield over 99.9% accuracy). DNA sequence analysis was facilitated by the PAP (version 2.4, D. Swofford) and MacClade.
Results and Discussion

Ribosomal RNA sequences have proved useful in discerning the relationships between distantly related taxa of eubacteria, archaeobacteria and eukaryotes (e.g. Woese, 1987). However, the slow rates of evolution of these molecules preclude their use in the fine-scale phylogenetic positioning of taxa; the 16S rRNA sequence differs by only 2–3% between E. coli and Salmonella typhimurium (C. Woese, personal communication). While substitutions at this locus provide a powerful tool for discerning the relationships between very distantly related taxa, they do not supply adequate information for the elucidation of relationships between more closely related taxa, e.g. members of the same genus.

In contrast, typical protein-encoding genes evolve at a relatively rapid rate. The synonymous sites (positions at which substitutions would not alter the amino acid composition of the protein) differ by 50–60% between E. coli and Sal. typhimurium (Ochman & Wilson, 1988; Sharp & Li, 1987a). Assuming a random accumulation of substitutions at these sites, this level of divergence limits their utility in phylogenetic reconstruction of higher taxa because most sites are assumed to have mutated at least once and would therefore contain ambiguous phylogenetic information.

The bacterial genome contains a class of highly expressed genes which experience selective pressures on codon choice, reflected in high codon usage bias (Sharp & Li, 1987b). Not surprisingly these genes evolve more slowly at synonymous sites (Sharp & Li, 1987a). The gap and ompA loci are highly expressed genes which exhibit a high degree of codon bias (Table 1). Between E. coli and Sal. typhimurium, these genes differ by 16 and 20% at synonymous sites, respectively. This level of nucleotide divergence allows an adequate estimation of the genetic relationships between closely related species, while minimizing convergent evolutionary events. Our data indicate that DNA sequences of slowly evolving genes such as gap and ompA may be useful in determining the taxonomic placement of novel isolates, as well as in resolving the phylogenetic relationships amongst established taxa.

The nucleotide sequences of gap and ompA are the first to be obtained from all the representative enteric species subsequent to the divergence of the genus Serratia. The aligned sequences of portions of the gap and ompA loci from eighteen strains of enteric bacteria are presented in

Table 2. Nucleotide sequence differences between strains within species

<table>
<thead>
<tr>
<th>Strain*</th>
<th>Changes</th>
<th>Changes</th>
<th>Deletions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Escherichia blattae 29907/33429</td>
<td>3</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>29907/33430</td>
<td>2</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>33429/33430</td>
<td>3</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Escherichia fergusonii 35469/35471</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>35469/35472</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>35471/35472</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Escherichia hermannii 33650/33652</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Escherichia vulneris 29943/33821</td>
<td>71</td>
<td>94</td>
<td>3</td>
</tr>
<tr>
<td>29943/33822</td>
<td>73</td>
<td>95</td>
<td>3</td>
</tr>
<tr>
<td>33821/33822</td>
<td>5</td>
<td>7</td>
<td>0</td>
</tr>
</tbody>
</table>

* ATCC numbers from Table 1.

Fig. 1. The ompA sequences from E. coli, Sal. typhimurium and Serratia marcescens and the gap sequence from E. coli agree well with previously determined sequences (Beck & Bremer, 1980; Branlant & Branlant, 1985; Braun & Cole, 1984; Freud1 & Cole, 1983). The G+C contents of the sequences range from 49 to 56 mol% (Table 1) and, for the most part, reflect the overall genomic G+C content of the species (Normore & Brown, 1970). However, the G+C contents of these loci from Serratia do not reflect the overall G+C content of the Serratia genome (59 mol%; Normore & Brown, 1970), indicating that the nucleotide content of these genes is highly conserved across enteric species and, as noted by Sharp (1990), the selective forces reflected in codon usage biases are not constrained by mutational pressures towards high G+C content. The sequences from each species display high levels of codon usage bias (Table 1), as calculated by the codon adaptation index (CAI) of Sharp & Li (1987b). Although the CAI is based on the pools of preferred codons estimated from the E. coli genome, these data suggest that preferred codons are probably similar among divergent enteric species, and are not significantly altered by selective constraints on overall genomic nucleotide content.

Table 2 shows that there is little sequence variation among conspecific strains. Only E. vulneris ATCC 29943 is very divergent from other strains typed to this species. This result is not unexpected, as total DNA relatedness is reported as being significantly lower between this and other putative E. vulneris isolates (Brenner et al., 1982b). Therefore, with the exception of two representative strains of E. vulneris (ATCC 29943 and ATCC 33821), DNA sequences from single strains of each species were
<table>
<thead>
<tr>
<th>Eoo</th>
<th>ATGGTTTC</th>
<th>CCGGTGTCT</th>
<th>72</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eb11</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Eb12</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Eb13</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ehe1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ehe2</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Evi1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Evi2</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Evi3</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Efe1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Efe2</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sty</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cfr</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Kpn</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Eae</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Smr</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sod</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

| Nucleotide sequences at the gap locus |
|---|---|---|
| Eco | GCGGTGTCT | 190 |
| Eb11 | A | - |
| Eb12 | A | - |
| Eb13 | A | - |
| Ehe1 | T | - |
| Ehe2 | T | - |
| Evi1 | T | - |
| Evi2 | T | - |
| Evi3 | T | - |
| Efe1 | T | - |
| Efe2 | T | - |
| Sty | T | - |
| Cfr | T | - |
| Kpn | T | - |
| Eae | A | - |
| Smr | A | - |
| Sod | A | - |

<table>
<thead>
<tr>
<th>Eco</th>
<th>GGTGGTTGC</th>
<th>288</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eb11</td>
<td>CA-C-C</td>
<td>A-T</td>
</tr>
<tr>
<td>Eb12</td>
<td>CA-C-C</td>
<td>A-T</td>
</tr>
<tr>
<td>Eb13</td>
<td>CA-C-C</td>
<td>A-T</td>
</tr>
<tr>
<td>Ehe1</td>
<td>G-C-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Ehe2</td>
<td>G-C-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Evi1</td>
<td>G-C-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Evi2</td>
<td>G-C-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Evi3</td>
<td>G-C-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Efe1</td>
<td>G-C-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Efe2</td>
<td>G-C-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Sty</td>
<td>G-C-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Cfr</td>
<td>G-C-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Kpn</td>
<td>G-C-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Eae</td>
<td>G-G-A-A-A</td>
<td>G-A</td>
</tr>
<tr>
<td>Sod</td>
<td>G-G-A-A-A</td>
<td>G-A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Eco</th>
<th>AAGCAACC</th>
<th>396</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eb11</td>
<td>G-C-C</td>
<td>C-G</td>
</tr>
<tr>
<td>Eb12</td>
<td>G-C-C</td>
<td>C-G</td>
</tr>
<tr>
<td>Eb13</td>
<td>G-C-C</td>
<td>C-G</td>
</tr>
<tr>
<td>Ehe1</td>
<td>TC-G-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Ehe2</td>
<td>TC-G-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Evi1</td>
<td>TC-G-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Evi2</td>
<td>TC-G-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Evi3</td>
<td>TC-G-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Efe1</td>
<td>TC-G-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Efe2</td>
<td>TC-G-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Sty</td>
<td>TC-G-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Cfr</td>
<td>TC-G-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Kpn</td>
<td>TC-G-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Eae</td>
<td>TC-G-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Smr</td>
<td>TC-G-G</td>
<td>T-C</td>
</tr>
<tr>
<td>Sod</td>
<td>TC-G-G</td>
<td>T-C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Eco</th>
<th>CTTGGTCA</th>
<th>423</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eb11</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Eb12</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Eb13</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ehe1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ehe2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Evi1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Evi2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Evi3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Efe1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Efe2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sty</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cfr</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Kpn</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Eae</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Smr</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sod</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Nucleotide sequences at the ompA locus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Eeo</strong></td>
<td><strong>Ehb</strong></td>
<td><strong>Eho</strong></td>
</tr>
<tr>
<td>CGCAACGC</td>
<td>ACTGACCG</td>
<td>TACACGCC</td>
</tr>
<tr>
<td>CGCAACGC</td>
<td>ACTGACCG</td>
<td>TACACGCC</td>
</tr>
<tr>
<td>CGCAACGC</td>
<td>ACTGACCG</td>
<td>TACACGCC</td>
</tr>
<tr>
<td>CGCAACGC</td>
<td>ACTGACCG</td>
<td>TACACGCC</td>
</tr>
<tr>
<td>CGCAACGC</td>
<td>ACTGACCG</td>
<td>TACACGCC</td>
</tr>
<tr>
<td>CGCAACGC</td>
<td>ACTGACCG</td>
<td>TACACGCC</td>
</tr>
<tr>
<td>CGCAACGC</td>
<td>ACTGACCG</td>
<td>TACACGCC</td>
</tr>
<tr>
<td>CGCAACGC</td>
<td>ACTGACCG</td>
<td>TACACGCC</td>
</tr>
<tr>
<td>CGCAACGC</td>
<td>ACTGACCG</td>
<td>TACACGCC</td>
</tr>
<tr>
<td>CGCAACGC</td>
<td>ACTGACCG</td>
<td>TACACGCC</td>
</tr>
<tr>
<td>CGCAACGC</td>
<td>ACTGACCG</td>
<td>TACACGCC</td>
</tr>
<tr>
<td>CGCAACGC</td>
<td>ACTGACCG</td>
<td>TACACGCC</td>
</tr>
<tr>
<td>CGCAACGC</td>
<td>ACTGACCG</td>
<td>TACACGCC</td>
</tr>
<tr>
<td>CGCAACGC</td>
<td>ACTGACCG</td>
<td>TACACGCC</td>
</tr>
<tr>
<td>CGCAACGC</td>
<td>ACTGACCG</td>
<td>TACACGCC</td>
</tr>
<tr>
<td>CGCAACGC</td>
<td>ACTGACCG</td>
<td>TACACGCC</td>
</tr>
<tr>
<td>CGCAACGC</td>
<td>ACTGACCG</td>
<td>TACACGCC</td>
</tr>
<tr>
<td>CGCAACGC</td>
<td>ACTGACCG</td>
<td>TACACGCC</td>
</tr>
</tbody>
</table>
Evolution of enteric bacteria

<table>
<thead>
<tr>
<th>630</th>
<th>657</th>
<th>684</th>
<th>711</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Eae</strong></td>
<td><strong>Eha</strong></td>
<td><strong>Ehe</strong></td>
<td><strong>Eha</strong></td>
</tr>
<tr>
<td><strong>Esa</strong></td>
<td><strong>Eae</strong></td>
<td><strong>Ehe</strong></td>
<td><strong>Ehe</strong></td>
</tr>
<tr>
<td><strong>Sod</strong></td>
<td><strong>Sod</strong></td>
<td><strong>Sod</strong></td>
<td><strong>Sod</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>738</th>
<th>765</th>
<th>792</th>
<th>819</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Eae</strong></td>
<td><strong>Eha</strong></td>
<td><strong>Ehe</strong></td>
<td><strong>Ehe</strong></td>
</tr>
<tr>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
</tr>
<tr>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
</tr>
<tr>
<td><strong>Sod</strong></td>
<td><strong>Sod</strong></td>
<td><strong>Sod</strong></td>
<td><strong>Sod</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>846</th>
<th>873</th>
<th>900</th>
<th>927</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
</tr>
<tr>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
</tr>
<tr>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
</tr>
<tr>
<td><strong>Sod</strong></td>
<td><strong>Sod</strong></td>
<td><strong>Sod</strong></td>
<td><strong>Sod</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>951</th>
<th>972</th>
<th>999</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
</tr>
<tr>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
</tr>
<tr>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
<td><strong>Ehr</strong></td>
</tr>
<tr>
<td><strong>Sod</strong></td>
<td><strong>Sod</strong></td>
<td><strong>Sod</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Eao</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CTGGATCAG</td>
<td>CTGTACAGC</td>
<td>CAGCTGAGC</td>
<td>AACTTGGAT</td>
</tr>
<tr>
<td>CCGAAAGAC</td>
<td>GGTTCCGTA</td>
<td>GTTGTTCTG</td>
<td>GGTTACACC</td>
</tr>
<tr>
<td>GACCGCATC</td>
<td>GCTTACAAC</td>
<td>CAGGGTCTG</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Ehr</th>
<th>Ehr</th>
<th>Ehr</th>
<th>Ehr</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sty</strong></td>
<td><strong>Cfr</strong></td>
<td><strong>Fpn</strong></td>
<td><strong>Eae</strong></td>
<td><strong>SmA</strong></td>
</tr>
<tr>
<td><strong>Sod</strong></td>
<td><strong>Sod</strong></td>
<td><strong>Sod</strong></td>
<td><strong>Sod</strong></td>
<td><strong>Sod</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Ehr</th>
<th>Ehr</th>
<th>Ehr</th>
<th>Ehr</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sty</strong></td>
<td><strong>Cfr</strong></td>
<td><strong>Fpn</strong></td>
<td><strong>Eae</strong></td>
<td><strong>SmA</strong></td>
</tr>
<tr>
<td><strong>Sod</strong></td>
<td><strong>Sod</strong></td>
<td><strong>Sod</strong></td>
<td><strong>Sod</strong></td>
<td><strong>Sod</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Eao</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CTCAGTACG</td>
<td>CTCAGTACG</td>
<td>CTCAGTACG</td>
<td>CTCAGTACG</td>
<td></td>
</tr>
<tr>
<td>GCGATCGAC</td>
<td>GCAGGCAGC</td>
<td>GCATCGGC</td>
<td>GATGCAGC</td>
<td></td>
</tr>
<tr>
<td>TCTGCTGCG</td>
<td>GAACTCGGC</td>
<td>GCATCGGC</td>
<td>GATGCAGC</td>
<td></td>
</tr>
<tr>
<td>ACTGACCG</td>
<td>GTGACGTC</td>
<td>GCATCGGC</td>
<td>GATGCAGC</td>
<td></td>
</tr>
<tr>
<td>TCTGCTGCG</td>
<td>GAACTCGGC</td>
<td>GCATCGGC</td>
<td>GATGCAGC</td>
<td></td>
</tr>
<tr>
<td>GCATCGGC</td>
<td>GATGCAGC</td>
<td>GCATCGGC</td>
<td>GATGCAGC</td>
<td></td>
</tr>
<tr>
<td>CTCAGTACG</td>
<td>CTCAGTACG</td>
<td>CTCAGTACG</td>
<td>CTCAGTACG</td>
<td></td>
</tr>
<tr>
<td>GCGATCGAC</td>
<td>GCAGGCAGC</td>
<td>GCATCGGC</td>
<td>GATGCAGC</td>
<td></td>
</tr>
<tr>
<td>TCTGCTGCG</td>
<td>GAACTCGGC</td>
<td>GCATCGGC</td>
<td>GATGCAGC</td>
<td></td>
</tr>
<tr>
<td>ACTGACCG</td>
<td>GTGACGTC</td>
<td>GCATCGGC</td>
<td>GATGCAGC</td>
<td></td>
</tr>
<tr>
<td>TCTGCTGCG</td>
<td>GAACTCGGC</td>
<td>GCATCGGC</td>
<td>GATGCAGC</td>
<td></td>
</tr>
<tr>
<td>GCATCGGC</td>
<td>GATGCAGC</td>
<td>GCATCGGC</td>
<td>GATGCAGC</td>
<td></td>
</tr>
<tr>
<td>CTCAGTACG</td>
<td>CTCAGTACG</td>
<td>CTCAGTACG</td>
<td>CTCAGTACG</td>
<td></td>
</tr>
<tr>
<td>GCGATCGAC</td>
<td>GCAGGCAGC</td>
<td>GCATCGGC</td>
<td>GATGCAGC</td>
<td></td>
</tr>
<tr>
<td>TCTGCTGCG</td>
<td>GAACTCGGC</td>
<td>GCATCGGC</td>
<td>GATGCAGC</td>
<td></td>
</tr>
<tr>
<td>ACTGACCG</td>
<td>GTGACGTC</td>
<td>GCATCGGC</td>
<td>GATGCAGC</td>
<td></td>
</tr>
</tbody>
</table>
analysed further. The phylogenetic relationships among isolates were analysed for each gene using a maximum parsimony algorithm (PAUP). In both cases, a single most parsimonious tree was found. Moreover, the branching orders for the gap and ompA loci were congruent. This common tree is presented in Fig. 2. The relative branching order of the species was tested using the Felsenstein test on four-taxon subgroups, assuming a stochastic molecular clock (Felsenstein, 1985). This test allows a statistical comparison of the three possible relationships between four taxa, and confirmed all branch points (the resolution of relationships between E. vulneris, E. hermannii and Enterobacter aerogenes required the elimination of synonymous changes from the data set).

Since the sequences of gap and ompA both define a single phylogenetic tree, it is likely that this genealogy accurately reflects the evolutionary relationships among bacterial lineages. More specifically, it is unlikely that horizontal transfer of both genes between enteric species (requiring independent transfer of the two loci or conjugative transfer of a portion of the bacterial chromosomes exceeding 1000 kb) has occurred. In addition, the sequences from independent isolates typed as the same species are quite similar, suggesting that the original classification of these strains within a species, based on total DNA relatedness and overall phenotypic similarity, is valid. The overall branching order and pattern of divergence among taxa (E. coli, Sal. typhimurium, Klebsiella pneumoniae, Ent. aerogenes, and Serr. marcescens) reflected by these genes is similar to earlier assessments of their genetic relationships based on a combination of phenotypic and genetic characters (Brenner & Falkow, 1971; Cocks & Wilson, 1972; Ochman & Wilson, 1987).

Species of Salmonella are traditionally considered to be those most closely related to E. coli. If Escherichia were considered to represent a monophyletic group, additional species of Escherichia should be more closely related to E. coli than is Salmonella. From the phylogeny presented in Fig. 2, it is apparent that the genus Escherichia is not a monophyletic group. That is, neither the ancestor of all Escherichia species, nor all its descendant taxa, belong to that genus. The Felsenstein statistics presented in Table 3 (Felsenstein, 1985) support the hypothesis that only E. coli and E. fergusonii are congeneric species and diverged following the split of Escherichia from Salmonella (P < 0.05). However, this hypothesis is rejected for E. blattae, E. hermannii, and both genotypes of E. vulneris (P < 0.05). The data suggest that these Escherichia species are more distantly related to E. coli than is Salmonella and should be placed in alternative genera. Since bacterial genera should reflect monophyletic groups, the putative species of Escherichia must be reclassified or this genus would encompass species of Salmonella, Citrobacter, Enterobacter and Klebsiella. In a similar manner, Serr. odorifera was classified according to phenotypic characters (Grimont et al., 1978). As is evident from Fig. 2, such classification does not allow definition of the genus Serratia as a monophyletic group.

The DNA hybridization data of Brenner and co-workers (Brenner et al., 1982a, b) also indicate that both E. hermannii and E. vulneris are distantly related to E. coli. However, neither species could be unambiguously classified into any one genus based on DNA hybridization studies. The DNA hybridization data are summarized in Table 4. In the cases of E. hermannii and E. vulneris, metabolic characteristics implied close relationships to E. coli and to one another. These species were classified in the genus Escherichia on the basis of metabolic profiles, despite the data from DNA hybridiz-

---

**Fig. 2.** Phylogeny of selected enteric bacteria based on DNA sequences from the gap and ompA loci, generated by a maximum parsimony algorithm. Taxon designations refer to strains listed in Table 1. The phylogeny was rooted using the nucleotide sequence of the gap locus from a strain of Morganella morganii (H. Ochman, unpublished results).
Evolution of enteric bacteria

Table 3. Summary of statistics from Felsenstein's test

<table>
<thead>
<tr>
<th>Potential phylogeny:</th>
<th>C</th>
<th>S</th>
<th>Eco</th>
<th>**</th>
<th>Sty</th>
<th>OUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. blattae, gap</td>
<td>11</td>
<td>2</td>
<td>26*</td>
<td>15*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OUT: K. pneumoniae</td>
<td>7</td>
<td>9</td>
<td>21*</td>
<td>12*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. blattae, ompA</td>
<td>14*</td>
<td>5</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OUT: Serr. marcescens</td>
<td>26*</td>
<td>6</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. fergusonii, gap</td>
<td>4</td>
<td>6</td>
<td>26*</td>
<td>20*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OUT: Ent. aerogenes</td>
<td>9</td>
<td>10</td>
<td>19*</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. fergusonii, ompA</td>
<td>3</td>
<td>6</td>
<td>36*</td>
<td>30*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OUT: Ent. aerogenes</td>
<td>8</td>
<td>10</td>
<td>18*</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. hermannii, gap</td>
<td>1</td>
<td>3</td>
<td>36*</td>
<td>33*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OUT: K. pneumoniae</td>
<td>6</td>
<td>2</td>
<td>30*</td>
<td>24*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Taxon designations for potential phylogenies are: Eco, Escherichia coli K12; Sty, Salmonella typhimurium LT2; OUT, outgroup specified for each test; **, test species specified. Statistic 'C' is the number of informative characters supporting each tree; statistic 'S' corresponds to the number of steps separating the best tree from the next-best tree. *, Measures which are statistically significant at $P < 0.05$. E. vulneris 'a' and 'b' are strains ATCC 29943 and ATCC 33821, respectively.

Evolutionary relationships among enteric bacteria have utilized a wide range of characters, including immunological relationships (Cocks & Wilson, 1972), metabolic pathways (Ahmad et al., 1990; Jensen, 1985), DNA hybridization (Brenner, 1978; Brenner & Falkow, 1971) and standard biochemical and enzymic profiles (Brenner, 1981, 1984; Edwards & Ewing, 1962). An underlying weakness of these methods is the inability to directly measure the genetic relationships between species. Alternatively, an assessment of homologous DNA sequences among species not only resolves fine-scale taxonomic relationships but the data allow an analysis of gene evolution within the context of the evolutionary history of the organisms. Although it has been argued that constraints on the evolution of protein and ribosomal RNA molecules preclude their use as taxonomic tools (Meyer et al., 1986), one can statistically test phylogenies inferred from these macromolecules.

From published sequence data, it is possible to design oligonucleotide primers that will allow PCR amplification of homologous genes from distantly related taxa. This approach has been applied to examine the
evolutionary relationships among metazoans based on cytochrome \( b \) oxidase genes of mitochondrial DNA (Kocher et al., 1989) and could be extended to include any group of organisms. The only requirement in using macromolecules to infer accurate phylogenies is that their sequences have sufficient variation to discern evolutionary relationships but are minimally affected by convergent events. We have shown that genetic loci experiencing selection reflected in codon usage bias satisfy these requirements.

Owing to its ubiquitous distribution and conservative evolution, GAPDH would serve as an ideal molecule for inferring fine-scale phylogenetic relationships among organisms not readily distinguished by rDNA sequences. Aside from the eighteen gap genes from enteric bacteria reported in this paper, the nucleotide sequences of GAPDH-encoding genes have been determined for some 40 taxa, including representatives from each urkingdom (Doolittle et al., 1990). Using a large set of GAPDH-encoding sequences, Doolittle et al. (1990) postulated that the high degree of similarity between the \( E. \ coli \) gap gene and eukaryotic GAPDH-encoding sequences resulted from a horizontal transfer event introducing a eukaryotic gap homologue into the bacterial genome (see also Martin & Cerff, 1986). Other eubacterial lineages, notably \( Bacillus \) \( stearothermophilus \), \( Thermus \) \( aquaticus \), \( Zymomonas \) \( mobilis \) and the cyanobacteria-derived chloroplasts, harbour gap genes distinct from those of \( E. \ coli \) and eukaryotes. The gap sequences presented in this paper clearly establish that any horizontal transfer event must have occurred prior to the divergence of the enteric bacteria. Phylogenies based on the gap and \( ompA \) sequences are congruent (Fig. 2), indicating that these genes were not subject to horizontal transfer processes. Moreover, a second potential GAPDH-encoding locus has been identified in \( E. \ coli \) (Alefounder & Perham, 1989) which is similar to the non-enteric prokaryotic gap genes (Doolittle et al., 1990). To identify the lineage accommodating the initial transfer event, one must determine the sequences of GAPDH-encoding genes from species having diverged from the \( E. \ coli \) lineage prior to \( Serr. \) \( odorifera \), yet following the divergence of the purple bacterium \( Z. \ mobilis \).

While DNA sequence data may be useful in the generic placement of novel species, a revision of the taxonomy of extant enteric species would cause considerable confusion. The long-standing practice of classifying aberrant bacterial strains as novel species without regard to evolutionary history (e.g. the classification of non-lactose-fermenting strains of \( E. \ coli \) in \( Shigella \)) has constrained the ability of classification to reflect phylogeny. The advent of PCR and rapid DNA sequencing has allowed the elucidation of fine-scale phylogenetic relationships among bacterial species, which is necessary to examine the evolution of their genomes.

We thank J. P. Carulli, R. F. DuBose, D. E. Krane and A. Larson for aid and helpful discussions, Carl Woese for sharing unpublished data, and the laboratories of Doug Berg and Staffan Normark for providing bacterial strains. This work was supported by grant GM 40322 (D. L. H.) and GM 40995 (H. O.) from the National Institute of Health.

### References


ALEFOUNDER, P. R. & PERHAM, R. N. (1989). Identification, molecular cloning and sequence analysis of a gene cluster encoding the Class II fructose 1,6-bisphosphate aldolase, 3-phosphoglycerate kinase and a second glyceraldehyde 3-phosphate dehydrogenase of \( Escherichia \) \( coli \). Molecular Microbiology 3, 723–732.


### Table 4. Comparison of DNA hybridization and nucleotide divergences of \( Escherichia \) \( coli \) and other enteric bacteria

<table>
<thead>
<tr>
<th>Strain</th>
<th>DNA hybridization</th>
<th>Total changes†</th>
<th>gap</th>
<th>ompA</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E. \ coli ) K-12</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( E. \ coli ) fergusoni ATCC 35469</td>
<td>60</td>
<td>43</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>( S. ) typhimurium LT2</td>
<td>45</td>
<td>55</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>Citrobacter freundii OS60</td>
<td>50</td>
<td>62</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>( E. \ coli ) ATCC 33650</td>
<td>31</td>
<td>81</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>( E. \ coli ) valnieri ATCC 33821</td>
<td>35</td>
<td>94</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>( K. ) pneumoniae LD199</td>
<td>38</td>
<td>71</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>Enterobacter aerogenes E482</td>
<td>37</td>
<td>107</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>( E. \ coli ) blattae ATCC 29907</td>
<td>29</td>
<td>75</td>
<td>104</td>
<td></td>
</tr>
<tr>
<td>Serratia marcescens ATCC 13880</td>
<td>24</td>
<td>128</td>
<td>145</td>
<td></td>
</tr>
</tbody>
</table>

* Percentage association of genomic DNA at 60 °C (Brenner & Falkow, 1971; Brenner et al., 1982a, b; Farmer et al., 1983).
† Total number of substitutions between \( E. \ coli \) and the given taxon at the respective locus.
‡ The Kendall rank test \( S \) statistic testing an inverse correlation between the total number of substitutions between \( E. \ coli \) and the other enteric strains at each locus and their percentage reassociation during DNA hybridization.
§ denotes a significant negative correlation \( (P < 0.05) \).


