Evolution of the gyrB gene and the molecular phylogeny of Enterobacteriaceae: a model molecule for molecular systematic studies

Catherine Dauga†

Phylogenetic trees showing the evolutionary relatedness of Enterobacteriaceae based upon gyrB and 16S rRNA genes were compared. Congruence among trees of these molecules indicates that the genomes of these species are not completely mosaic and that molecular systematic studies can be carried out. Phylogenetic trees based on gyrB sequences appeared to be more reliable at determining relationships among Serratia species than trees based on 16S rRNA gene sequences. gyrB sequences from Serratia species formed a monophyletic group validated by significant bootstrap values. Serratia fonticola had the most deeply branching gyrB sequence in the Serratia monophyletic group, which was consistent with its atypical phenotypic characteristics. Klebsiella and Enterobacter genera seemed to be polyphyletic, but the branching patterns of gyrB and 16S rRNA gene trees were not congruent. Enterobacter aerogenes was grouped with Klebsiella pneumoniae on the gyrB phylogenetic tree, which supports that this species could be transferred to the Klebsiella genus. Unfortunately, 16S rRNA and gyrB phylogenetic trees gave conflicting evolutionary relationships for Citrobacter freundii because of its unusual gyrB evolutionary process. gyrB lateral gene transfer was suspected for Hafnia alvei. Saturation of gyrB genes was observed by the pairwise comparison of Proteus spp., Providencia alcalifaciens and Morganella morganii sequences. Depending on their level of variability, 16S rRNA gene sequences were useful for describing phylogenetic relationships between distantly related Enterobacteriaceae, whereas gyrB sequence comparison was useful for inferring intra- and some intergeneric relationships.

Keywords: gyrB, 16S rRNA, Enterobacteriaceae, evolutionary systematic studies

INTRODUCTION

The enteric bacteria belonging to the Enterobacteriaceae family are of special microbiological interest because of their pathogenic and non-pathogenic relationships with the human gastrointestinal tract. Some Enterobacteriaceae are also plant or animal pathogens and are responsible for substantial economic losses in agriculture (Brenner, 1981). This high diversity of hosts correlates with a wide phenotypic diversity. There are currently over 35 genera described in this family (Brenner, 1984; Farmer et al., 1985; Hickman-Brenner et al., 1985; Gherna, 1991; Hauben et al., 1998).

When investigating the environmental distribution of Enterobacteriaceae, it must be remembered that it is not always possible to culture natural populations of bacteria by standard techniques (Amann et al., 1995). They may be viable but non-culturable cells because their physiological capacities have been modified. This has been described for Photobacterium (Xenorhabdus) luminescens, Escherichia coli O157:H7 and Salmonella typhi (Morgan et al., 1997; Tanaka et al., 2000; Cho & Kim, 1999). Evolutionary systematics which classify bacteria naturally can be used to identify these difficult

† Present address: Collection de l’Institut Pasteur, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France.

Abbreviations: GCIII, third codon position G+C; RSCU, relative synonymous codon usage.

The EMBL accession numbers of the gyrB sequences are AJ300528–AJ300554.
to culture organisms (Debavov, 1999; Dore et al., 1998; Vuddhakul et al., 2000). However, lineages must be placed in their correct evolutionary position before gene sequence analysis of clades can reliably identify organisms (Ludwig et al., 1998).

Phylogenetic studies on Enterobacteriaceae and other prokaryotes have used rRNA sequences (Fox, 1980; Olsen et al., 1986; Woese, 1987). Comparison of the 16S rRNA gene sequences of Serratia, Yersinia, Salmonella, Photobacterium and Erwinia have demonstrated that this molecule can discriminate between closely related species (Dauga et al., 1990; Ibrahim et al., 1993; Christensen et al., 1998; Liu et al., 1997). The phylogenetic analysis of 16S rRNA gene sequences showed four emerging clusters of phytopathogens within the Enterobacteriaceae (Hauben et al., 1998).

However, unlike DNA hybridization studies (Brenner, 1984) and phenotypic classification, 16S rRNA studies found that Klebsiella, Enterobacter and Serratia genera are polyphyletic (Hauben et al., 1998; Spröer et al., 1999).

Protein-encoding gene sequences have also been used to describe the phylogenetic relationships of Enterobacteriaceae. Unfortunately, in spite of a higher phylogenetic resolution than that obtained from 16S rRNA sequences (Yamamoto & Harayama, 1996), the phylogenetic trees obtained were not totally consistent with the classical taxonomy. The sequences of gap genes, which encode glyceraldehyde-3-phosphate dehydrogenase, and ompA, which encodes outer-membrane protein 3A, revealed that five species of Escherichia do not form a monophyletic group (Lawrence et al., 1991). The sequences of recA, a gene involved in recombination and DNA repair, showed that Shigella strains cannot be phylogenetically separated from Escherichia coli (Lloyd & Sharp, 1993). The groE sequences, which encode stress proteins, indicated that the members of genera Enterobacter, Pantoea and Klebsiella were closely related to each other and did not form distinct phylogenetic groups (Harada & Ishikawa, 1997). The partial sequence of rpoB, the gene encoding the RNA polymerase β-subunit, and the partial sequence of infB, the gene encoding translation initiation factor 2, also suggested that Klebsiella is a polyphyletic genus (Mollet et al., 1997; Hedegaard et al., 1999).

Recently, the gyrB gene was proposed as a suitable phylogenetic marker for the identification and classification of bacteria (Yamamoto & Harayama, 1996; Yamamoto et al., 1999). gyrB is a single-copy gene, present in all bacteria, which encodes the ATPase domain of DNA gyrase, an enzyme essential for DNA replication (Huang, 1996). The amino acid sequences of Gyrb are conservative enough to allow the comparison of taxa which are not closely related (Yamamoto & Harayama, 1996; Huang, 1996). The phylogenetic analyses of gyrB nucleotide sequences also reflected the evolutionary relationships of closely related species such as in the Acinetobacter (Yamamoto et al., 1999), Pseudomonas (Yamamoto & Harayama, 1998), Shevannella (Venkateswaran et al., 1999) and Micromonaspora (Kasai et al., 2000) genera. In addition, gyrB sequence analyses also distinguished Pseudomonas putida strains (Yamamoto & Harayama, 1995), slowly growing mycobacterial species (Kasai et al., 2000), Vibrio parahaemolyticus from Vibrio alginolyticus (Venkateswaran et al., 1998), and Bacillus cereus from Bacillus mycoides, Bacillus thuringiensis and Bacillus anthracis (Yamada et al., 1999).

In this study I investigated the reliability and usefulness of gyrB gene sequences for the phylogenetic analysis of Enterobacteriaceae. The gyrB-based phylogenetic trees were compared to 16S rRNA gene (16S rDNA) sequence trees for the whole Serratia genus, two species each of Proteus, Enterobacter and Klebsiella and seven other species belonging to seven other genera of Enterobacteriaceae.

gyrB partial sequence analysis proved to be more reliable and useful than 16S rDNA for determining the evolutionary relationships of Serratia species. However, many discrepancies occurred between the classic descriptions of genera and monophyletic groups obtained with gyrB phylogenetic trees for Enterobacter and Klebsiella species. Therefore, (i) I determined whether saturation and/or long-branch-attracting phenomena distorted the gyrB phylogenetic trees, and (ii) codon usage and base composition analyses of genes were used to determine whether genetic transfer occurred in some lineages. Knowledge of some of these phylogenetic construction biases and natural events allowed me to determine when the evolution of the gyrB genes best reflected the phylogenetic relationships of species within the Enterobacteriaceae.

METHODS

Bacterial strains. The strains used in this study are listed in Table 1.

16S rRNA data. Several 16S rRNA sequences are available in GenBank and EMBL databases: Citrobacter freundii (accession no. M59291; unpublished), Escherichia coli (X80725; Cilia et al., 1996), Enterobacter aerogenes (AB004750; Harada et al., 1996), Enterobacter cloacae (AJ251469; unpublished), Pectobacterium carotovorum (Erwinia carotovora) (M59149; unpublished), Hafnia alvei (M59155; unpublished), Haemophilus influenzae (M53019 and M599433; unpublished), Klebsiella pneumoniae (X87276; Ludwig et al., 1995), Klebsiella terrigena (Y17658; unpublished), Plesiononas shigelloides (M59159; unpublished) and Salmonella typhimurium (X80681; Cilia et al., 1996). Nine Serratia spp. 16S rRNA sequences were also recovered from international databases (AJ233427, AJ233428, AJ233429, AJ233430, AJ233431, AJ233432, AJ233433, AJ233434, AJ233436; Spröer et al., 1999). Recently, I published the 16S rRNA gene sequences from Proteus mirabilis (AJ301682), Proteus vulgaris (AJ301683), Morganella morganii (AJ301681) and Providencia alcalifaciens (AJ301684) in the EMBL database.

Chromosomal DNA isolation. Each bacterial sample was grown aerobically in Trypto-casein soy broth (Sanofi Diag-
Table 1. Strains and gyrB sequences used in this study

<table>
<thead>
<tr>
<th>Species</th>
<th>Strain*</th>
<th>Characteristics</th>
<th>Accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citrobacter freundii</td>
<td>CIP 57.32⁹ = ATCC 8090⁹</td>
<td></td>
<td>AJ300528</td>
</tr>
<tr>
<td>Enterobacter aerogenes</td>
<td>CIP 60.86⁹ = ATCC 13048⁹</td>
<td></td>
<td>AJ300554</td>
</tr>
<tr>
<td>Enterobacter cloacae</td>
<td>CIP 60.85⁹ = ATCC 13047⁹</td>
<td></td>
<td>AJ300553</td>
</tr>
<tr>
<td>Haemilia alvei</td>
<td>CIP 57.31⁹ = ATCC 13337⁹</td>
<td>Biotype 1</td>
<td>AJ300551</td>
</tr>
<tr>
<td>Klebsiella pneumoniae</td>
<td>CIP 82.91⁹ = ATCC 13883⁹</td>
<td>Subsp. pneumoniae</td>
<td>AJ300550</td>
</tr>
<tr>
<td>Klebsiella terrigena</td>
<td>CIP 80.7⁹ = ATCC 33257⁹</td>
<td></td>
<td>AJ300549</td>
</tr>
<tr>
<td>Morganella morganii</td>
<td>CIP A231⁹ = ATCC 25830⁹</td>
<td></td>
<td>AJ300548</td>
</tr>
<tr>
<td>Pectobacterium carotovorum</td>
<td>CIP 82.83⁹ = ATCC 15713⁹</td>
<td>Subsp. carotovorum</td>
<td>AJ300552</td>
</tr>
<tr>
<td>Plesiomonas shigelloides</td>
<td>CIP 63.5⁹ = ATCC 14029⁹</td>
<td></td>
<td>AJ300545</td>
</tr>
<tr>
<td>Proteus mirabilis</td>
<td>CIP 103181⁹ = ATCC 29906⁹</td>
<td></td>
<td>AJ300546</td>
</tr>
<tr>
<td>Proteus vulgaris</td>
<td>CIP 58.60⁹ = ATCC 13315⁹</td>
<td></td>
<td>AJ300544</td>
</tr>
<tr>
<td>Providencia alcalifaciens</td>
<td>CIP 82.90⁹ = ATCC 9868⁹</td>
<td></td>
<td>AJ300547</td>
</tr>
<tr>
<td>Salmonella enterica</td>
<td>CIP 60.62⁹ = ATCC 43971⁹</td>
<td>Subsp. enterica serotype Typhimurium</td>
<td>AJ300529</td>
</tr>
<tr>
<td>Serratia entomophila</td>
<td>CIP 102919⁹ = ATCC 43705⁹</td>
<td>Biotype 1</td>
<td>AJ300543</td>
</tr>
<tr>
<td>Serratia entomophila</td>
<td>A15†</td>
<td>Biotype 2</td>
<td>AJ300542</td>
</tr>
<tr>
<td>Serratia ficaria</td>
<td>CIP 79.23⁹ = ATCC 33105⁹</td>
<td>O-type 1</td>
<td>AJ300541</td>
</tr>
<tr>
<td>Serratia ficaria</td>
<td>G4037†</td>
<td>O-type 2</td>
<td>AJ300540</td>
</tr>
<tr>
<td>Serratia fonticola</td>
<td>CIP 78.64⁹ = ATCC 29844⁹</td>
<td></td>
<td>AJ300539</td>
</tr>
<tr>
<td>Serratia grimesii</td>
<td>CIP 103361⁹ = ATCC 14460⁹</td>
<td></td>
<td>AJ300538</td>
</tr>
<tr>
<td>Serratia liquefaciens</td>
<td>CIP 103238⁹ = ATCC 27592⁹</td>
<td></td>
<td>AJ300537</td>
</tr>
<tr>
<td>Serratia marcescens</td>
<td>CIP 103235⁹ = ATCC 13880⁹</td>
<td>Biotype A2a</td>
<td>AJ300536</td>
</tr>
<tr>
<td>Serratia marcescens</td>
<td>G3770†</td>
<td>Biotype A6a</td>
<td>AJ300535</td>
</tr>
<tr>
<td>Serratia marcescens</td>
<td>G81†</td>
<td>Biotype TCT</td>
<td>AJ300534</td>
</tr>
<tr>
<td>Serratia odorifera</td>
<td>CIP 79.1⁹ = ATCC 33077⁹</td>
<td></td>
<td>AJ300533</td>
</tr>
<tr>
<td>Serratia plymuthica</td>
<td>CIP 103239⁹ = ATCC 183⁹</td>
<td></td>
<td>AJ300532</td>
</tr>
<tr>
<td>Serratia proteamaculans</td>
<td>CIP 103236⁹ = ATCC 19323⁹</td>
<td>Biotype Clc</td>
<td>AJ300531</td>
</tr>
<tr>
<td>Serratia rubidea</td>
<td>CIP 103234⁹ = ATCC 27593⁹</td>
<td></td>
<td>AJ300530</td>
</tr>
</tbody>
</table>

* CIP, Collection Institut Pasteur, Paris, France; ATCC, American Type Culture Collection, Manassas, VA, USA.
† From the collection of P. A. D. Grimont, Institut Pasteur, Paris, France.

**GyrB gene amplification.** The following primers were used to amplify gyrB: gyr-320, 5'-TAARTTyGAyGAYAaCT-CyTAYAAaGt-3' (R = A or G; Y = C or T); rgyr-1260, 5'-CMCCyTCCaCCTaGATTC-GTCC-3' (M = A or C). Primer gyr-320 annealed at positions 306-332 and primer rgyr-1260 annealed at positions 1276-1255 (Escherichia coli numbering; Adachi et al., 1987). These primers were selected on the basis of gyrB sequence conservation in Escherichia coli (accession numbers X04341 and X00870; Adachi et al., 1987). Proteus mirabilis (M58352 and M31295; Skovgaard, 1990), Pseudomonas putida (D37926 and X5461; Yamamoto & Hayaraya, 1995; Paraske & Harwood, 1990) and Haemophilus influenzae (L45208 and L42023; Fleischmann et al., 1995) from the γ subclass of the Proteobacteria. The gyr-320 primer annealed at the ATP binding site of the gyrB subunit (Wang, 1996). The rgyr-1260 primer annealed to part of the topoisomerase II region used to define the consensus pattern [(LIVMA)-x-E-G, which recognizes the DNA topoisomerase II family in the PROSITE database (Bairoch et al., 1997). Primer specificity was confirmed that the selected primers should not amplify topoisomerase IV sequences or other coding sequences. The reaction mixture (100 μl) contained 10 mM Tris/HCl (pH 8.3), 50 mM KCl, 2.5 mM MgCl₂, 0.1% gelatin, 0.2 mM each dNTP (Pharmacia Biotech), 50 pmol each primer, 1.25 U Hi-Tag polymerase (Bioprobe) and 1 μl DNA. The mixture was overlaid with 50 μl mineral oil (Sigma). PCR amplification was carried out in a PTC-100 thermal Cycler (MJ Research) as follows: 94 °C for 4 min, followed by 35 cycles of 94 °C for 1 min, 55 °C for 1 min and 72 °C for 2 min, with a final incubation at 72 °C for 10 min. Amplification products were examined by standard agarose gel electrophoresis (0.8%) and ethidium bromide staining. A 1 kb DNA ladder (Gibco-BRL) was used as a molecular size marker.

**GyrB gene cloning and sequencing.** To facilitate cloning a 12-base dUMP-containing sequence (CUACUACUACUA) was added to the 5' end of PCR primers (Rashchian et al., 1992). The CLONAMP System (Gibco-BRL) was used to clone the resulting PCR product into pAMP1, and this was subsequently used to transform Escherichia coli MC1061. The presence of the insert in five clones was checked by restriction analysis with BamHI and EcoRI (Amersham International). Two clones per sample were sequenced by use of the thermosequenase core sequencing kit, with 7-
deaza-dGTP and universal primers complementary to plasmid, on a Vistra DNA Sequencer (Amersham). The following primers were used for sequencing: gyr-550, S'-GAATTCGARTACGATATTYTGG-3', and rgyr-550 which annealed at positions 541–562; gyr-670S, S'-AAGGGGGCGYRTCMRKGC-3' (K = G or T), and rgyr-670S at positions 656–671; gyr-670P, S'-GAAGGYGGTATYMACGC-3' and rgyr-670P at positions 655–674; gyr-860, S'-CATKGCSSYRSGGAARCC-3' (S = G or C), and rgyr-860 at positions 850–867; and gyr-1010, S'-AAGACA-AACTGTTTCTTCCGA-3', and rgyr-1010 at positions 1010–1031 (Escherichia coli numbering system). The EMBL accession numbers of the gyrB sequences determined are given in Table 1.

Phylogenetic data analysis. The CLUSTAL V algorithm (Higgins, 1994) of the MegAlign program in the Lasergene software package (DNASTar, Madison, WI) was used to make initial alignments with default gap penalties. The initial alignments were further refined by eye, introducing gaps to improve the overall alignment, and by use of the alignment editor of the PAUP* version 4.0 software package (Swofford, 1998). Sequence distance matrices were established in pairwise comparisons by use of the algorithm (Kimura, 1980). Phylogenetic trees were constructed by the neighbour-joining method (Saitou & Nei, 1987) using the PHYLIP version 3.5 software package (Felsenstein, 1993). Parsimony analysis was carried out with maximum-parsimony implemented in the PAUP* 4.0 software package. Maximum-parsimony trees were obtained by 100 random addition heuristic search replicates and the tree bisection-reconnection branch-swapping option. A maximum-likelihood analysis was also done with PAUP* version 4.0 with a transition/transversion ratio of 1:5 for both 16S rDNA and gyrB genes with and without a gamma distribution for rate variation among sites (Felsenstein, 1981; Olsen et al., 1994). Statistical significance was evaluated by bootstrap analysis (Felsenstein, 1985) with 100 repeats of bootstrap samplings.

Saturation test. Saturation was detected graphically by plotting sequence differences between pairs of taxa against the estimated number of substitutions between sequences (Griffiths, 1997). The sequence differences between pairs of taxa corresponded to the observed proportion of positions that are not shared by two taxa (uncorrected distance). The estimated numbers of substitutions per site between all pairs of taxa on trees used in this study were derived from the parsimony criterion. These trees were rooted by using the method of Nei & Gojobori (1986) in the PAML package (Yang, 1997).

RESULTS

16S rRNA-based phylogenetic tree. Evolutionary trees, based on 16S rDNA sequences, were constructed by the neighbour-joining, parsimony and maximum-likelihood methods. The phylogenetic trees were based on 1546 unambiguously aligned positions, 161 of which were informative under the parsimony criterion. These trees were rooted by using Plesiomonas shigelloides, which is the most closely related species to the Enterobacteriaceae family (Brenner, 1981). The dendrogram in Fig. 1 shows the phylogeny of Enterobacteriaceae obtained by the distance method.

As expected, 16S rDNA sequences from Proteus mirabilis and Proteus vulgaris were closely clustered, regardless the phylogenetic method used. This monophyletic cluster was found to be grouped with sequences from Providencia alcalifaciens and Morganella morganii, which are also closely phenotypically related to each other (kauffmann, 1954). Unfortunately, the phylogenetic position of this cluster, rooted by Hafnia alvei, differed according to the phylogenetic method used. It was found to be a sister group of Pectobacterium carotenovorum, Escherichia coli, Salmonella enterica, Enterobacter spp., Citrobacter freundii and Klebsiella spp. sequences on the distance tree, but was within one of the Serratia species clusters on the parsimony and maximum-likelihood trees (Fig. 1).

16S rDNA sequences from Pectobacterium carotenovorum, Escherichia coli, Salmonella enterica, Enterobacter spp., Citrobacter freundii and Klebsiella spp. formed a monophyletic group regardless of the method used. However, the relationships between sequences from Enterobacter spp., Citrobacter freundii and Klebsiella spp. overlapped and were inconsistent with...
bacterial taxonomy. Enterobacter cloacae was more closely related to Salmonella enterica and Citrobacter freundii, whereas Enterobacter aerogenes grouped with Klebsiella terrigena (Fig. 1).

Two separate clusters (no significant bootstrap values) were obtained for Serratia species, as described by Spröer et al. (1999). One cluster contained Serratia marcescens, Serratia rubidaea, Serratia odorifera, Serratia ficaria and Serratia entomophila. Serratia ficaria and Serratia entomophila formed the deepest branch of the distance tree (Fig. 1). The second cluster, which contained Serratia plymuthica, Serratia grimesii and Serratia proteamaculans, also included Serratia fonticola (Dauga et al., 1990; Spröer et al., 1999) which does not have all the key phenotypic characteristics of the Serratia genus (Gavini et al., 1979).

16S rDNA sequence comparisons did not confirm the current groups of Serratia, Klebsiella and Enterobacter species (Hauben et al., 1998; Spröer et al., 1999). The bootstrap value at the nodes of a few clusters obtained were too low to induce much confidence. 16S rDNA sequences alone could not resolve the phylogenetic relationships between the Enterobacteriaceae species used. The similarity of the nine Serratia species examined was between 93.9 and 99.3%, and for species from different genera of Enterobacteriaceae the similarity was between 88.1 and 98.1%. This low rate of variation of 16S rDNA explains the poor phylogenetic information obtained from their sequences. Therefore, I studied gyrB which has a more variable nucleotide sequence.

**Nucleotide polymorphism of gyrB**

The nucleotide sequence of a 971 bp segment of the gyrB gene encoding 323 aa was determined for two strains of Serratia ficaria species, two strains of Serratia entomophila, three strains of Serratia marcescens and one strain for each of the other seven species of Serratia. gyrB genes were also sequenced for 13 strains belonging to 13 species from 10 different genera of Enterobacteriaceae and for the Plesiomonas shigelloides type strain. It was found that amino acids 103, 109, 335 and 337 [Escherichia coli amino acid numbering system (X04341/X00870)], corresponding to sequence base positions 1–3, 21–23, 699–701 and 705–707, and amino acids 114–126, corresponding to sequence base positions 36–72, encoding a glycine-rich helix, were conserved. Both of these regions are involved in maintaining contact with ATP (Huang, 1996). As expected, amino acids Arg-136 (base positions 102–105) and Gly-164 (186–188) (Huang, 1996; Wang, 1996), which are the targets of coumarin- and novobiocin-type antibiotics, were found in all of the sequences. I identified 119 polymorphic nucleotide sites (occurring in 20 codons) in the variable regions of Klebsiella sequences, 120 polymorphic nucleotide sites...
Fig. 2. Phylogenetic trees derived from gyrB sequence comparisons with neighbour-joining (a), parsimony (b) and maximum-likelihood (c) methods. The scale bar represents the number of substitutions per nucleotide position (a and c). Percentages refer to significant bootstrap values of 100 calculated trees (a and b).

(occurring in 18 codons) in Enterobacter sequences, 104 polymorphic nucleotide sites (occurring in 11 codons) in Proteus sequences and 273 polymorphic nucleotide sites (occurring in 48 codons) in the 10 Serratia sequences. One of these sites harboured a Serratia signature sequence. In all Serratia species, gyrB sequences encoded a lysine (K) at codon 206 (base positions 312–314) in a β-sheet-shaped region of the ATP binding site.

The similarity between nucleotide sequences from three biotypes of Serratia marcescens (Table 1) ranged from 96.5 to 98.9%. Nucleotide sequences from two biotypes of Serratia ficaria and two biotypes of Serratia entomophila showed 99.4 and 99.5% similarities, respectively. These variations were attributable to intraspecies differences. The nucleotide variation of Serratia marcescens strains was greater than that of Serratia ficaria or Serratia entomophila, although the amount of amino acid variation was almost identical in the three species. gyrB nucleotide sequence variability allowed closely related strains to be distinguished within these three Serratia species.

The similarity between gyrB nucleotide sequences from all Serratia species ranged from 84.8 to 97.3%. The similarity between their amino acid sequences ranged from 92 to 99.4%. Nucleotide sequence similarities within the genera Klebsiella, Enterobacter and Proteus were 87.1, 87.4 and 88.9%, respectively, and their amino acid sequence similarities were 93.8, 94.4 and 96.6%, respectively. The amount of variation within Klebsiella, Enterobacter and Proteus were similar to those for Serratia. The gyrB nucleotide sequences contained a mean of three times more mutations than 16S rDNA sequences. The high variability of gyrB nucleotide sequences might provide more phylogenetic information than 16S rDNA and enable species relationships within the Enterobacteriaceae to be resolved.

gyrB-based phylogenetic tree

The phylogenetic trees based on gyrB nucleic acid sequences and obtained with neighbour-joining, parsimony and maximum-likelihood methods are shown in Fig. 2(a), (b) and (c), respectively. They were based on 971 unambiguously aligned positions, corresponding to 381 informative positions under the conditions of parsimony. As before, the trees were rooted by using the Plesiomonas shigelloides sequence.
Molecular phylogeny of Enterobacteriaceae

Regardless of the phylogenetic methods employed, the Proteus, Providencia and Morganella sequences formed a monophyletic cluster validated with a significant bootstrap value branching out the other enteric species (Fig. 2a, b and c). gyrB sequences from Escherichia, Salmonella, Klebsiella, Enterobacter and Citrobacter species also formed a monophyletic group with a significant bootstrap value (Fig. 2a and b). Serratia gyrB sequences formed one monophyletic group, which was validated by a significant bootstrap value (96%) on the distance phylogenetic tree (Fig. 2a), and supported by a bootstrap value of 75% on the maximum-parsimony phylogenetic tree (Fig. 2b). The phylogenetic positions of Pectobacterium carotovorum and Hafnia alvei remained variable according to the phylogenetic method used (Fig. 2a, b and c).

Two phylogenetic clusters, validated with significant bootstrap values, were found within the Serratia group. The first cluster included Serratia entomophila and Serratia ficaria, from insects, and Serratia marcescens, a human opportunistic pathogen. The second cluster grouped Serratia grimesii with Serratia liquefaciens, Serratia proteamaculans and Serratia plymuthica, three different species initially included on the basis of phenotypic characteristics in the species Serratia liquefaciens (Grimont et al., 1982a, b). Consistent with its atypical phenotypic characteristics, Serratia fonticola had the most deeply branching gyrB sequence in the Serratia group (Fig. 2a, b and c).

Both Klebsiella and Enterobacter genera remained polyphyletic on the gyrB phylogenetic tree. gyrB sequences from Klebsiella pneumoniae and Klebsiella terrigena, which are often confused by commercial identification systems (Monnet & Freney, 1994), did not place these two species in the same group. Likewise, sequences from Enterobacter cloacae and Enterobacter aerogenes, two important hospital pathogens, were not grouped on the tree. In fact, Enterobacter aerogenes was more closely related to Klebsiella pneumoniae. This relationship is consistent with the small number of

Fig. 3. Neighbour-joining phylogenetic tree obtained from GyrB protein sequences. The scale bar represents the estimated number of amino acid replacements per position. Percentages refer to significant bootstrap values of 100 calculated trees. The broken lines with an arrowhead show the position of the species pair Serratia odorifera and Serratia fonticola and the Salmonella enterica-Citrobacter freundii-Escherichia coli-Enterobacter spp.-Klebsiella spp. cluster obtained with maximum-parsimony method.

http://ijs.sgmjournals.org 537
The phylogenetic relationships determined by gyrB genes and GyrB protein sequence analyses were more consistent than those established by the 16S rRNA phylogenetic tree for Klebsiella pneumoniae and Enterobacter aerogenes. Because of its high variability, gyrB gene phylogenetic analysis was more reliable than 16S rRNA and GyrB protein analyses to infer intrageneric relationships of Serratia species. However, although the gyrB phylogenetic tree contained 13 strongly supported groups, the use of different phylogenetic methods revealed uncertainties in the branching order of some other Enterobacteriaceae. Therefore, I tried to determine whether phylogenetic construction bias, such as saturation or long-branch-attracting phenomena, could explain the lack of resolution of gyrB genes for inferring some phylogenetic relationships.

Limitation of gyrB-gene-based phylogeny due to high evolutionary rates

Due to the high evolutionary rate of the gyrB gene, multiple substitutions often occur at the same nucleotide position. This implies that there has been mutational saturation of gyrB nucleotide sequences, which is known to confuse phylogenetic inferences (Leblond-Bourget et al., 1996). This phenomenon was obvious when the Haemophilus influenzae sequence was added. This resulted in the Enterobacteriaceae family cluster being broken (data not shown). Thus, the high levels of gyrB nucleotide sequence variation did not allow me to determine the relationships between distantly related organisms.

I compared inferred substitutions with observed differences in the gyrB sequences of Enterobacteriaceae (data not shown) to determine whether multiple changes that accumulated in the gyrB sequences resulted in false identities and masked the actual number of evolutionary events (Philippe et al., 1994; Griffiths, 1997; Roe et al., 1997). Saturation analysis clearly indicated that Providencia alcalafaciensis, Morganella morganii, Proteus mirabilis and Proteus vulgaris sequences were mutationally saturated.

In addition, these species were too deeply rooted in the gyrB phylogenetic tree attracted by outgroup (Fig. 2a, b and c). This phenomenon is known as long branch attracting (Philippe & Laurent, 1998) and occurred because the evolutionary rates of this branch were three times greater than in other branches.

Saturation and long branch attracting both meant that the phylogenetic positions obtained with gyrB genes of Providencia alcalafaciensis, Morganella morganii, Proteus mirabilis and Proteus vulgaris were uncertain.

Nucleotide composition, codon usage and the genetic transfer hypothesis

Horizontal gene transfer may explain some of the discrepancies between the defined clusters on phylogenetic gene trees and the groups predicted by phenotypic data. Horizontally transferred genes reflect the genome composition of the donor at the time of introgression, and over time they acquire the DNA composition (biases and asymmetries) of the new genome (Lawrence & Ochman, 1997). Therefore, the G+C content [total gene G+C content and third codon position G+C content (GCIII)] and the relatedness of codon usage of gyrB genes were studied.

The variations between the total G+C contents of the gyrB genes and the mean genomic G+C content of the strains used in this study were not significantly large.
enough to clearly indicate interspecies gene transfer. The total G + C content of the gyrB genes was similar or slightly higher than the genomic G + C content for each species (Table 2). As the divergence in the G + C content of genes may be related to their chromosomal location (Deschavanne & Filipski, 1995; Lobry, 1996) the G + C contents of gyrB genes were compared with those of genes from the same genome region (when available). The genome data from Escherichia coli, Salmonella enterica, Serratia marcescens and Proteus mirabilis (Burland et al., 1993; Sanderson et al., 1995; Skovgaard & Hansen, 1987; Skovgaard, 1990) showed that the total G + C contents of gyrB genes were approximately equal to those of the neighbouring genes: recF (required for recombination and repair), dnaN (encoding the DNA polymerase III β-subunit) and dnaA (required for initiation of chromosomal replication). No proof of lateral gene transfer was obtained for these species.

The highest GCIII content values correlated with high values of genomic G + C content for Serratia rubidaea, Serratia marcescens, Serratia ficaria and Serratia entomophila (Table 2). The GCIII content of gyrB sequences was lower than the low overall genomic G + C content of Proteus mirabilis, Proteus vulgaris and Providencia alcalifaciens. The GCIII content was generally higher than the overall genomic G + C content in genomes with medium G + C content, as described by Muto & Osawa (1987).

The large differences observed between the total G + C content and GCIII content of gyrB sequences arose because the third codon position has less functional importance (Muto & Osawa, 1987; Majumdar et al., 1999), except for Hafnia alvei in which GCIII content was equal to the total G + C content of gyrB. This atypical GCIII content may have arisen due to the ancient acquisition of a foreign gene.

I also sought a codon usage bias among gyrB genes to detect gene transfer. Genes in species that are closely phylogenetically related tended to be rather homogeneous in codon usage. Closely related species, such as Escherichia coli and Salmonella enterica, have very similar codon preferences, whereas distant species, such as Bacillus subtilis, have different preferential codon usage patterns (Andersson & Kurland, 1990). Therefore, I compared codon usages of gyrB genes among Enterobacteriaceae species.

---

**Table 2. Comparison between gyrB, recF, dnaN and dnaA genes and genome nucleotide compositions**

<table>
<thead>
<tr>
<th></th>
<th>G + C nucleotide composition (mol%)</th>
<th>Genomic G + C content (mol%)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>gyrB</td>
<td>gyrB (at 3rd codon)</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>54.94</td>
<td>63.47</td>
</tr>
<tr>
<td>Citrobacter freundii</td>
<td>53.91</td>
<td>59.75</td>
</tr>
<tr>
<td><em>Salmonella enterica</em> ser. Typhimurium</td>
<td>55.04</td>
<td>65.58</td>
</tr>
<tr>
<td>Enterobacter aerogenes</td>
<td>55.56</td>
<td>65.94</td>
</tr>
<tr>
<td>Enterobacter cloacae</td>
<td>54.22</td>
<td>65.01</td>
</tr>
<tr>
<td>Pectobacterium carotovorum</td>
<td>53.23</td>
<td>59.56</td>
</tr>
<tr>
<td>Hafnia alvei</td>
<td>49.38</td>
<td>49.22</td>
</tr>
<tr>
<td>Klebsiella pneumoniae</td>
<td>57.10</td>
<td>69.65</td>
</tr>
<tr>
<td>Klebsiella terrigena</td>
<td>55.97</td>
<td>67.49</td>
</tr>
<tr>
<td>Serratia entomophila</td>
<td>59.05</td>
<td>77.10</td>
</tr>
<tr>
<td>Serratia ficaria</td>
<td>59.36</td>
<td>78.63</td>
</tr>
<tr>
<td>Serratia fonticola</td>
<td>54.32</td>
<td>65.63</td>
</tr>
<tr>
<td>Serratia grimesii</td>
<td>52.67</td>
<td>59.75</td>
</tr>
<tr>
<td>Serratia liquefaciens</td>
<td>55.25</td>
<td>67.49</td>
</tr>
<tr>
<td>Serratia marcescens</td>
<td>59.77</td>
<td>77.40</td>
</tr>
<tr>
<td>Serratia odorifera</td>
<td>57.51</td>
<td>74.60</td>
</tr>
<tr>
<td>Serratia plymuthica</td>
<td>56.07</td>
<td>69.04</td>
</tr>
<tr>
<td>Serratia proteamaculans</td>
<td>52.67</td>
<td>61.00</td>
</tr>
<tr>
<td>Serratia rubidaea</td>
<td>60.39</td>
<td>78.94</td>
</tr>
<tr>
<td>Proteus mirabilis</td>
<td>41.67</td>
<td>29.72</td>
</tr>
<tr>
<td>Proteus vulgaris</td>
<td>40.64</td>
<td>29.10</td>
</tr>
<tr>
<td>Providencia alcalifaciens</td>
<td>44.65</td>
<td>37.77</td>
</tr>
<tr>
<td>Morganella morgani</td>
<td>53.19</td>
<td>60.68</td>
</tr>
<tr>
<td>Plesiomonas shigelloides</td>
<td>54.18</td>
<td>61.60</td>
</tr>
</tbody>
</table>

* Genomic G + C content (at the species level) from Krieg & Holt (1984).
† Genomic G + C content from Serratia entomophila strain ATCC 43705T (Grimont et al., 1988).
‡ Mean genomic G + C content of the Serratia liquefaciens species complex (Grimont et al., 1982b).
The analysis was carried out on the RSCU data. Correspondence analysis of the results (Fig. 4) identified the major trends in codon usage: the vertical axis is associated with GCIII (Musto et al., 1998), the horizontal axis is correlated with the frequencies of codons ending in C or U versus A or G (Fennoy & Bailey-Serres, 1993). The codon choice of the Escherichia coli gyrB gene corresponds to the optimal codon choice described for highly expressed genes (Ikemura, 1985; Bulmer, 1988). The codon usages of gyrB genes of Serratia spp., Klebsiella spp., Enterobacter spp., Pectobacterium carotovorum, Citrobacter freundii and Salmonella enterica were almost comparable to that of Escherichia coli on the RSCU correspondence analysis plot (Fig. 4). Analysis of codon usage did not show any bias that might be the result of gene transfer in these species.

The codon usage of Proteus mirabilis, Proteus vulgaris and Providencia alcalifaciens is most characteristic of A + T-biased micro-organisms in which codons with an A or T in the third position are used preferentially (Ohtaka & Ishikawa, 1993).

The differences between Escherichia coli and Hafnia alvei codon usage are principally due to the use of U instead of C at the first or the third codon positions for the amino acids Asp, Gly, His, Leu, Pro and Ser. The atypical codon usage of the gyrB gene was confirmed when the codon usages of the genes encoding the RNA polymerase β-subunit (rpoB, accession number U77438) and translational initiation factor 2 (infB, accession number AJ227978) of the Hafnia alvei genome were checked (data not shown). This extremely atypical codon usage (different from that of Proteus spp.) and the unusual gyrB GCIII content may reflect the acquisition of gyrB by horizontal transfer in Hafnia alvei.

### Table 3. Number of substitutions per synonymous site (d_s) and per non-synonymous site (d_N) calculated by the method of Nei & Gojobori (1986)

Statistical significance between means of d_N or d_s/d_N from nucleotide positions 1–200 and 201–971 were calculated by t-tests with P values of 5 (*) and 1% (**). Values that differ between nucleotide positions 1–200 and 201–971 are in bold type.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Domain I (1–200)</th>
<th>Domain II (201–971)</th>
<th>(1–971)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>d_N</td>
<td>d_s</td>
<td>d_s/d_N</td>
</tr>
<tr>
<td><strong>Within:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serratia genus</td>
<td>0.045</td>
<td>0.562</td>
<td>0.085</td>
</tr>
<tr>
<td>Serratia liquefaciens-Serratia plymuthica-</td>
<td>0.012</td>
<td>0.520</td>
<td>0.025</td>
</tr>
<tr>
<td>Serratia grimesii-Serratia proteamaculans</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serratia entomophila-Serratia ficaria-Serratia marcescens</td>
<td>0.021</td>
<td>0.318</td>
<td>0.070</td>
</tr>
<tr>
<td>Enterobacter genus</td>
<td>0.085</td>
<td>0.564</td>
<td>0.151</td>
</tr>
<tr>
<td>Klebsiella genus</td>
<td>0.134</td>
<td>0.603</td>
<td>0.223</td>
</tr>
<tr>
<td><strong>Between:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enterobacter aerogenes and Klebsiella pneumoniae</td>
<td>0.028</td>
<td>0.366</td>
<td>0.078</td>
</tr>
<tr>
<td>Enterobacter cloacae and Escherichia coli</td>
<td>0.021</td>
<td>0.670</td>
<td>0.032</td>
</tr>
<tr>
<td>Escherichia coli and Salmonella enterica</td>
<td>0.021</td>
<td>0.771</td>
<td>0.028</td>
</tr>
<tr>
<td>Citrobacter freundii and Escherichia coli</td>
<td>0.081</td>
<td>1.287</td>
<td>0.063</td>
</tr>
<tr>
<td>Citrobacter freundii and Salmonella enterica</td>
<td>0.105</td>
<td>1.699</td>
<td>0.062</td>
</tr>
<tr>
<td>Citrobacter freundii and Klebsiella terrigena</td>
<td>0.058</td>
<td>0.643</td>
<td>0.090</td>
</tr>
</tbody>
</table>
Molecular phylogeny of Enterobacteriaceae

**Synonymous and non-synonymous substitution analysis**

Mutation rates may vary among sites because of different selective constraints at different sites depending on the functional and/or structural requirements of the gene or protein. Failure to account for rate variation can lead to incorrect construction of phylogenies (Yang, 1996). Therefore, I analysed the number of synonymous substitutions per synonymous site ($d_S$) and the number of non-synonymous substitutions per non-synonymous site ($d_N$) for the entire nucleotide sequences. Visual analysis of the alignment showed that amino acid substitutions occurred more frequently at the 5' end of the sequence, thus $d_S$ and $d_N$ were also calculated separately for the first 200 nt (domain I) and for the next 771 nt (domain II) (Table 3).

$d_S$ is lower than $d_N$ in all cases (Table 3). Therefore, there is clear evidence of selective constraint on amino acid replacements. gyrB has evolved in all the phylogenetic branches under negative or purifying selection.

Analysis of $d_S$ sometimes revealed significant differences between sequence domains from species of the same genus. On average, $d_S$ for sites in domain I was approximately four times greater than for sites in domain II for *Serratia entomophila*, *Serratia ficaria* *Serratia marcescens* and *Enterobacter* species. $d_S$ was 15 times greater in domain I than in domain II for *Klebsiella* species. This comparison was the highest found for any pairwise comparison, which suggests that the amount of purifying selection on domain I differs according to the *Klebsiella* species.

When *Citrobacter freundii* was compared to *Salmonella enterica* or *Escherichia coli*, $d_S$ was approximately four times greater for sites in domain I than for sites in domain II. This surprising result was also obtained with three other *Citrobacter* sp. sequences (accession numbers AF005699, AF005700 and AF005701) (data not shown). This accelerated rate of substitution had a relatively low impact on $d_N$ in each case and led to seven convergent amino acids, Arg-33, Glu-35, Gly-36, Val-38, Glu-40, Thr-42 and Val-44, between GyrB sequences from *Citrobacter* spp. and *Klebsiella terrigena*. However, it did appear to create bias for the phylogenetic tree constructions based on both GyrB nucleic acid and GyrB amino acid sequences (Fig. 2 and 3). A phylogenetic tree constructed without sequences from domain I placed *Citrobacter freundii* in one monophyletic group with *Escherichia coli* and *Salmonella enterica* in agreement with phenotypic characters (Fig. 5). *Klebsiella terrigena*, *Klebsiella*
sequences. Related bacterial species than trees based on 16S rDNA reliable at determining relationships between closely rRNA and and their evolutionary relatedness based upon 16S DISCUSSION the sequences due to the unusual evolution of domain I of the Citrobacter freundii gyrB sequence.

The maximum-likelihood model with gamma rate was a good model of tree construction when high variation of substitution among sites occurred (Tateno et al., 1994). In our study, the gamma distribution of among-site variation rate had a shape parameter alpha of 0.82, meaning that most sites have low substitution rates, while a few sites have very high substitution rates (Yang, 1996). The position of Enterobacter cloacae was closer to the taxonomy on the maximum-likelihood tree with gamma rate (Fig. 2c) than without gamma rate (data not shown). Unfortunately, this model did not correct the position of Citrobacter freundii (Fig. 2c).

DISCUSSION

The phylogenetic position of some Enterobacteriaceae and their evolutionary relatedness based upon 16S rRNA and gyrB genes were compared.

Phylogenetic trees based on gyrB appeared to be more reliable at determining relationships between closely related bacterial species than trees based on 16S rDNA sequences. Serratia species formed a monophyletic group or clade based on gyrB sequences. This relationship is supported by high bootstrap values (96% for the distance gyrB phylogenetic tree and 75% for the maximum-parsimony gyrB phylogenetic tree). Thus I consider that a monophyletic relationship between Serratia species is very likely. This shows that phenotypic characteristics were a judicious choice of shared derived characters for defining the members of this genus. Evolutionary relationships between Serratia species were consistent with DNA relatedness clusters (Grimont et al., 1982b). The more distant position of Serratia fonticola in the gyrB phylogenetic tree matches the very low DNA–DNA hybridization values obtained between this species and the other Serratia species. In addition, the gyrB phylogenetic tree highlighted clades of Serratia with different pathogenic behaviour. Serratia entomophila, which is a pathogen of insects (Grimont et al., 1988), Serratia ficaria, an occasional pathogen in humans (Anahory et al., 1998), and Serratia marcescens, which causes nosocomial infections (Hejazi & Falkiner, 1997), formed a cluster. S rubidaceae, an opportunistic pathogen (Ursua et al., 1996), rooted this cluster (Fig. 2a, b and c). Of the environmental species (Serratia grimesii, Serratia plymuthica, Serratia liquefaciens and Serratia proteamaculans) only Serratia liquefaciens is an opportunistic pathogen, involved in hospital-acquired infections, especially in frail patients in which immune function is reduced. Serratia liquefaciens has also been identified as a cause of certain skin lesions (Nitzan et al., 1999). Serratia rubidaceae, Serratia marcescens, Serratia entomophila and Serratia ficaria might share virulence factors that could be inherited from their common ancestor, whereas if Serratia liquefaciens contains these virulence factors it did so by lateral gene transfer.

Phylogenetic trees based on gyrB gene sequences suggested that Klebsiella and Enterobacter species belonged to multiple phylogenetic groups. This confirms previous studies based on genes encoding translation initiation factor 2 (infB; Hedegaard et al., 1999), the RNA polymerase β-subunit (rpoB; Mollet et al., 1997), stress protein (groE; Harada & Ishikawa, 1997) and 16S rRNA (Hauben et al., 1998). Thus, Enterobacter and Klebsiella genera are validly described taxa, which harbour a collection of phylogenetically diverse species. Biochemical and metabolic characteristics might be unable to describe reliable hierarchical grouping of these genera in a cladistic manner. Indeed, phenotypes have a multifactorial genetic basis and their phenotypes could evolve convergently and confuse phenotypic cluster analysis (Lawrence et al., 1991). However, to confirm the trend observed with the phylogenetic analysis of gyrB genes, many more Enterobacter and Klebsiella species must be included.

In this study, the gyrB- and 16S-rDNA-based trees are not highly congruent for branching patterns of Enterobacter spp. and Klebsiella spp. gyrB and 16S rDNA sequences from Enterobacter aerogenes, Klebsiella terrigena and Klebsiella pneumoniae predicted different highly robust clusters for each molecular-based tree. The 16S rDNA phylogenetic tree grouped Enterobacter aerogenes with Klebsiella terrigena, even though they have different phenotypic characteristics (Izard et al., 1981). This unexpected clustering may be due to rRNA lateral gene transfer, but horizontal transfer was infrequent due to stringent functional constraints (Yap et al., 1999). 16S rDNA sequences are more likely to contain confused molecular synapomorphies and homoplasies, leading to incorrect phylogenetic constructions and high bootstrap values (Lecointre et al., 1993). Molecular phylogeny based on gyrB gene and protein sequences proposed that Enterobacter aerogenes clustered with Klebsiella pneumoniae, which is consistent with the phylogenetic tree based on groE gene sequences (Harada & Ishikawa, 1997). This is also consistent with the DNA hybridization data and proposals to transfer Enterobacter aerogenes to the Klebsiella genus (Brenner, 1984). The gyrB gene and GyrB protein phylogenetic analyses appeared to be better at determining relationships in this case than 16S rDNA.

The DNA–DNA hybridization method, which is the reference method for grouping conspecific bacterial strains (Wayne et al., 1987), has been used to determine the relationships between genera in the Enterobacteriaceae family (Brenner, 1984). However, noise, such as rate variation (with homoplasy), label compression and random experimental error can compromise the distance properties of DNA hybridization values.

C. Dauga
Providencia alcalifaciens may create long branch at-phylogenetic position of the phylogenetic position of this group. Although the Van de Peer & De Wachter, 1996), did not modify the correct both of these artifacts (Galtier & Gouy, 1995; gamma rate method, a phylogenetic method which can (Graybeal, 1998). The maximum-likelihood with samples used may have led to a biased placement utionary patterns and the low number of species (Lloyd & Sharp, 1993), there is a clearly different preferred codon usage for gyrB in these species. Thus, phylogenetic groupings based on nucleotide sequences may be directly affected by compositional bias of the genome, irrespective of the evolutionary history of the organisms (Hasegawa & Hashimoto, 1993). In addition, the high rate of sequence change for Proteus spp., Morganella morganii and Providencia alcalifaciens may create long branch attraction. Indeed, this combination of unusual evolutionary patterns and the low number of species samples used may have led to a biased placement (Graybeal, 1998). The maximum-likelihood with gamma rate method, a phylogenetic method which can correct both of these artifacts (Galtier & Gouy, 1995; Van de Peer & De Wachter, 1996), did not modify the phylogenetic position of this group. Although the phylogenetic position of the Proteus spp., Morganella morganii and Providencia alcalifaciens group was identical in the phylogenetic trees based on gyrB, 16S rRNA (Fig. 1), RNA polymerase β-subunit (rpoB) and translation initiation factor 2 (infB) sequences (Mollet et al., 1997; Hedegaard et al., 1999), the phylogenetic position of gyrB sequences for this group should be interpreted with caution, because of the possibility of artificial phylogenetic constructions.

To resolve problems due to the high evolutionary rates of nucleic acid data, phylogenetic analysis based on protein-encoding genes can be performed by using the amino acid sequences. Indeed, GyrB protein sequences showed no saturation of phylogenetic information within the Enterobacteriaceae family (data not shown), but their low contents in informative positions did not allow reliable study of phylogenetic relationships within the Serratia genus and between Enterobacter and Klebsiella species.

Discrepancies between the position of Enterobacter cloacae and Citrobacter freundii in gyrB- and 16S-rDNA-based phylogenetic trees suggest gyrB gene lateral transfer. However, codon usage bias and differences between the G + C content of the gene and the entire genome did not detect any genes acquired from closely related bacteria with similar base composition or codon usage. Slight variations of codon usage and G + C contents of gyrB genes in Escherichia coli, Salmonella enterica, Citrobacter freundii, Pectobacterium carotovorum, Enterobacter spp., Klebsiella spp. and Serratia spp. were not significant enough to confirm lateral gene transfer between these species.

Slight variations in the G + C content of gyrB genes were probably not due to different gene locations in relation to the origin of replication (Jensen et al., 1999; Karlin et al., 1998). Similarly, slight variations in codon usage in gyrB genes were not the result of different gene orientations on the bacterial chromosome due to strand compositional asymmetry (McLean et al., 1998). The chromosome maps and recent genomic data from Escherichia coli, Salmonella enterica, Klebsiella pneumoniae and Proteus mirabilis (Skovgaard, 1990), as well as Yersinia pestis (http:// genome.wustl.edu/gsc/Projects/bacterial/egadsinfo. shtml), Pseudomonas putida (Fujita et al., 1989), Haemophilus influenzae (Fleischmann et al., 1995) and Bacillus subtilis (Fujita et al., 1989), show that the genome region around gyrB is highly conserved. gyrB is located in a chromosomal region that has a central role in the initiation of chromosomal replication. This region is conserved in many Eubacteria and was probably inherited from the ancestral bacterium (Fujita et al., 1989). Therefore, the gyrB gene is probably rarely transferred and is thus a good tool for determining the history of Eubacteria.

The phylogenetic position of Hafnia alvei in the 16S-rRNA-based tree (Fig. 1) and the gyrB-based tree (Fig. 2a, b and Fig. 5) seemed almost congruent. However, the low number of informative positions of sequences for both molecules led to different tree topologies according to the phylogenetic tree methods used. No evidence for lateral gene transfer of 16S rRNA gene segments was obtained because of the lack of a large number of non-random base variations in the alignment of the stable region of the gene (data not shown) (Wang & Zhang, 2000). On the other hand, the obvious atypical codon usage and unusual GCIII contents in Hafnia alvei suggested that foreign gyrB gene transfer had occurred in this species. In this case, the phylogenetic position of Hafnia alvei could not be determined, because gyrB trees accurately reflect the evolutionary history of the gene, but not the history of the species (Lawrence, 1999).

The branching of Citrobacter freundii on the gyrB phylogenetic tree was related to the number of synonymous substitutions, reflecting the rate of neutral substitutions (Lenski et al., 1989), observed in different domains of the molecule. Citrobacter freundii, which had significantly different numbers of synonymous substitutions between domains compared to Escherichia coli and Salmonella enterica, branched in the vicinity of Klebsiella terrigena. The phylogenetic method of maximum-likelihood with gamma rate could not correct this, probably because it was too
Most of the differences between the gyrB and 16S rDNA trees are in areas of low resolution (differences generated by the different methods) or low bootstrap values for one or both of the molecules. The low resolution of the deep branches in the gyrB trees may have been due to poor representation of certain taxa, such as Pectobacterium carotovorum. Future studies will include more species to enable us to assess the overall phylogenetic relationships of Enterobacteriaceae. The number of sequences available at the ICB database (http://www.mbio.co.jp) is growing rapidly. Therefore, gyrB genes could prove useful for evolutionary systematic studies.

Differences in the resolution and the significance of particular branches helped me to define which molecules were useful for specific levels of phylogenetic comparisons: 16S rRNA phylogenetic studies, the reference method for defining phylogenetic lineages (Olsen & Woese, 1993), described relationships between distantly related Enterobacteriaceae, whereas the gyrB sequence was more appropriate for inferring intra- and some intergeneric relationships. In the future, 16S rDNA combined with the gyrB sequence method may be able to clarify molecular relationships of endosymbiotic Enterobacteriaceae (Heddi et al., 1998; Chen et al., 1999; Sameshima et al., 1999) and to identify viable but non-culturable Enterobacteriaceae in the environment (Vuddhakul et al., 2000).

ACKNOWLEDGEMENTS

Thanks are due to C. Bizet for providing type strains. Thanks are due to P. A. D. Grimont for providing biotypes of Serratia marcescens, Serratia ficaria and Serratia entomoaphila. Thanks are also due to S. Yamamoto and P. Bouvet for helpful discussions.

REFERENCES


