Patterns of sequence variation in two regions of the 16S rRNA multigene family of Escherichia coli

A. J. Martínez-Murcia,† A. I. Antón and F. Rodríguez-Valera

Sequence heterogeneities of variable positions located at regions V1 and V6 of 56 cloned 16S rRNA genes were determined from six Escherichia coli strains. These nucleotides were involved in secondary structure base-pairing of stem-loops. Compensatory and single mutations have occurred but secondary structure was conserved. Eight different sequences were found in the stem at region V1 indicating that in these sites mutation rates are higher than those of homogenization processes. Region V6 showed two different structures (V6-I and V6-II) although heterogeneities were determined in nine sites. Strains ECOR52 and ECOR56 only showed the V6-I sequence, ECOR35 showed V6-II, whereas clones from ECOR42 and ECOR49 showed both types of V6 structures. Results were confirmed by PCR using V6 sequence-specific probes. Stem V6-II was also found in 16S rRNA sequences deposited in the RDP (Ribosomal Database Project) belonging to distantly related taxa; ancestral sequence V6-II seems to be homogenized in all rrr operons of the multigene family of strain ECOR35 producing effects of distortion in the molecular clock, similar to those that homoplasies could produce. V6 sequence-specific probes were applied to the 72 ECOR strains: half showed both V6-I and V6-II, and the rest had one or another. Only strain ECOR24 did not yield products in the PCR test and sequencing of 12 cloned 16S rRNA genes revealed a third form, V6-III, also found in the RDP. Concerted evolution by homogenization of the rRNA family may induce chronometric distortions responsible for a loss of ultrametricity in phylogenetic trees, particularly, of very closely related micro-organisms.

Keywords: 16S rRNA, multigene family, Escherichia coli

INTRODUCTION

The sequencing of nucleic acids has provided a revolutionary approach to measuring evolutionary relationships. It is generally considered that the majority of genotypic change appears to be selectively neutral (Kimura, 1983), because genotypic variety is continually rolling, producing a profusion of change that happens regardless of whether the phenotype (function) is altered as a consequence. This neutrality contributes to the quasi-random nature of genotypic change, and it is the main reason biologists refer to sequence change as ‘clock-like’ (Olsen & Woese, 1993). Since the number of possible functional configurations for a given gene is so high, it is assumed that sequence similarity can never reflect convergent evolution (Woese, 1987).

Small-subunit rRNAs (16S rRNA in prokaryotes) are currently the most widely used informational macromolecules for tracing evolution (Gutell et al., 1994; Olsen & Woese, 1993; Olsen et al., 1994; Woese, 1987). They show a high degree of functional constancy, different positions in their sequences change at very different rates and they are universally spread. Microbial phylogenies inferred from rRNAs, RNA polymerases, elongation factor G, proton-translocating ATPases and several other genes give similar reconstructions of the past (Cammarano et al., 1992;
showed a tree with very uneven tips. For phylogenetic analysis, considering that molecular evolution is divergent and constant in rate, the comparison of properly aligned sequences leads ideally to phylogenetic trees whose tips lie at the same level, as in phenograms. Such relationships are described as ultrametric (Sneath, 1993). When a phylogenetic tree lacks ultrametricity, it shows very uneven tips suggesting chronometric distortions such as inconstant evolution rates and parallel or convergent evolutionary changes (Martínez-Murcia, 1993; Sneath, 1993). Hybrid events, involving genetic crossing-over of rrn from different individuals, have been considered on the basis of that evolution of different parts (domains) of the molecule have not been concordant (Sneath, 1993). The study of Cilia et al. (1996) shows a non-ultrametric tree and horizontal transfer was suggested to explain the sequence patterns found in variable regions of the rrn operons from a single strain.

In the present study, 56 cloned 16S rRNA genes derived from six *E. coli* strains selected from the ECOR reference collection were subjected to sequencing of the rrn operons in this multigene family and the distribution of sequence signatures among a reference population of this species. The information gained may contribute to an understanding of the mode of molecular evolution in *E. coli*.

**METHODS**

**Bacterial strains and cultivation.** *E. coli* strains of the ECOR reference collection includes 72 isolates from a wide variety of hosts and geographic regions (Ochman & Selander, 1984) and were kindly supplied by R. K. Selander. Their relationships have been determined by multilocus enzyme electrophoresis (MLEE) (Herzer, 1990) and by sequencing of the 16S–23S rRNA intergenic spacer region (García-Martínez et al., 1996). Duplicates of ECOR35, ECOR49 and ECOR52

<table>
<thead>
<tr>
<th>Primer/probe</th>
<th>Sequence (5' → 3')</th>
<th>Position*</th>
</tr>
</thead>
<tbody>
<tr>
<td>16S0F</td>
<td>AGAGTTTGATCATGTCGTCAG</td>
<td>8–27</td>
</tr>
<tr>
<td>16S5R</td>
<td>TTACCAGGCTGCTGCGACG</td>
<td>533–514</td>
</tr>
<tr>
<td>16S9R</td>
<td>CCGTCAATTCTTGGTGTGTTT</td>
<td>926–908</td>
</tr>
<tr>
<td>16S11F</td>
<td>TCCCCGAACGAGGCGAACAC</td>
<td>1096–1114</td>
</tr>
<tr>
<td>16S12R</td>
<td>CATTGTAGACGACGCTGAGCC</td>
<td>1241–1225</td>
</tr>
<tr>
<td>16S15R</td>
<td>GGTACCTTTGTTGACCTTT</td>
<td>1492–1510</td>
</tr>
<tr>
<td>V6-I</td>
<td>ACAATTCGATCTTGGAAAG</td>
<td>1006–1025</td>
</tr>
<tr>
<td>V6-II</td>
<td>ACCAATCCGATCTTGGAAAG</td>
<td>1006–1025</td>
</tr>
<tr>
<td>V6-III</td>
<td>ACCCTCGTATTCTCAAGAA</td>
<td>1006–1025</td>
</tr>
</tbody>
</table>

* Following the *E. coli* numbering system (Brosius et al., 1978).
were obtained from the American Type Culture Collection (ATCC 35354, ATCC 35368 and ATCC 35371, respectively). Bacteria were grown in Luria–Bertani medium for 18 h at 37 °C.

**DNA extraction.** Bacterial colonies were harvested in 200 μl of InstaGene Purification Matrix (Bio-Rad) by centrifuging at 10000 r.p.m. for 1 min in a Biofuge 13 (Heraeus Sputech) and the resultant pellet was incubated at 56 °C for 20 min. The tube was vortexed at high speed for 10 s, placed in a boiling water bath for 8 min and vortexed again at high speed for 10 s. After centrifuging at 12000 r.p.m. for 2 min the supernatant was removed and stored at -20 °C. DNA concentrations were spectrophotometrically estimated with the GeneQuant RNA-DNA calculator (Pharmacia LKB Biotechnology) in 0.5 mm diameter quartz capillaries. DNA dilutions of 20 ng μl⁻¹ were used for PCR.

**Oligonucleotide primers.** Characteristics of primers for PCR and sequencing of the 16S rRNA genes and probes V6-I, V6-II and V6-III are listed in Table 1. Oligonucleotides were purchased from Pharmacia Biotech.

**PCR amplification.** Approximately 40 ng chromosomal DNA was subjected to PCR amplification in a total volume of 50 μl containing 50 mM KCl, 10 mM Tris/HCl (pH 9.0), 1.5 mM MgCl₂, 0.1% Triton X-100 and 0.2 mM each deoxyribonucleotide (dATP, dCTP, dGTP, dTTP; Pharmacia LKB Biotechnology), 2 U TaqI DNA polymerase (Promega) and 50–60 pmol each primer (16SOF and 16S15R). The reaction mixtures were subjected to a thermal cycling regime (35 cycles on a ‘Peltier-effect’ thermal cycler PTC-100; MJ Research) of 94 °C for 30 s, 55 °C for 15 s and 72 °C for 2 min; a final single step of 72 °C for 5 min was then done. For probes V6-I, V6-II and V6-III (combined with the primer 16SOF), the same conditions were used except for an annealing step of 64 and 67 °C, respectively. Negative controls with no addition of template DNA were included. Following amplification, 5 μl of the PCR products were electrophoresed on 1% agarose gels in TAE buffer. The 1 Kb DNA ladder (Gibco BRL) was used as molecular marker. Agarose gels were stained with 0.5 μg ethidium bromide ml⁻¹, visualized with an ultraviolet lamp (λ = 312 nm) and photographed with Polaroid type 665 films.

**Fig. 1.** Secondary structures of a stem–loop at region V1 of the 16S rRNA genes from strains of E. coli. Numbers indicate positions as by Brosius et al. (1978). Variable nucleotides are in bold. Continuous arrows are between structures that differed by compensatory mutations (boxes); broken arrows are where single mutations appeared (circles). The number of times that each structure has been determined is in parentheses.
Purification and cloning of the PCR products. PCR products were purified using the GeneClean II kit (Bio 101), incubated with 2 U Klenow polymerase at 37 °C for 15 min, and purified with MicroSpin S-300 columns (Pharmacia). DNA fragments were inserted into the Smal site of the pUC-18 vector (Pharmacia Biotech) by ligation at 16 °C overnight. Competent DH5α Escherichia coli was transformed according to a standard protocol (Sambrook et al., 1989). Plasmid DNA was extracted using the Magic Miniprep DNA Purification System (Promega).

Sequencing of 16S rDNA. Sequences from two separately cloned fragments amplified in different PCR experiments were determined for each strain by the dideoxy-chain termination method (Sanger et al., 1977) using the Cy5 AutoRead Sequencing kit (Pharmacia). The sequencing products were separated on the ALFexpress DNA sequencer (Pharmacia) following the manufacturer's instructions.

Data analysis. Sequences from E. coli strains were manually aligned to a 16S rRNA previously published (Brosius et al., 1978). Secondary structure of stem–loops with variable sequence was constructed following the secondary structure from Stern et al. (1988). Comparative analysis with 16S rRNA sequences of the Ribosomal Database Project (RDP) was carried out by sending sequences to the programs CHECK-PROBE and SIMILARITY-RANK (Maidak et al., 1994).

RESULTS

Products of 16S rDNA amplification from five E. coli strains were cloned, almost complete 16S rRNA gene sequences of two clones of both ECOR35 and ECOR52, and one clone of ECOR49 were determined, aligned and variable positions were located. Sequences of variable V1 and V6 regions [nomenclature according to Huysmans & De Wachter (1986)] were primarily determined in 44 randomly selected clones: nine of ECOR56, 10 of ECOR49, 13 of ECOR35, and six of both ECOR52 and ECOR42. The clones for each strain were obtained from different PCR experiments. Secondary structures of the two stem–loops [positions 73–97 and 1000–1040; E. coli numbering according to Brosius et al. (1978)] were constructed following the secondary structure from Stern et al. (1988). Nucleotide heterogeneities determined in these stems (so-called stem V1 and stem V6) are shown in Figs 1 and 2. Seven different sequences were found in stem V1 (types I, IIA, IIB, IIIA, IIIB, IIIIC and IV), all of which conserve secondary structure base-pairing. Stem V6 was of two types (I and II); numbers of canonical and non-canonical pairs at positions where nucleotides change were the same in both stems, V6-I and V6-II.

Sequence diversity of stem V1

Six nucleotide positions appeared to be highly variable in stem V1, although mutations were not at random (Fig. 1). Composition change at stem V1 was: i, compensatory mutations, a change at one position is compensated by a change at its pairing position (pairs at positions 79–90 and 80–89), considered dependent changes as they have a structural relationship; and ii, single mutations in a non-canonical pair (positions 76–93) giving rise to a canonical pair or vice versa. Base composition at positions 79–90 were: pair G–C in V1-I and V1-II (A and B); pair U–A in V1-III (A, B and C); and pair A–U in V1-IV. At positions 80–89, pair C–G was found in all sequences of stem V1 except for V1-I that showed A–U. Theoretical structure V1-X (not yet determined) was proposed as an intermediate between V1-IIA or V1-IIIA and V1-IV; it shows a compensatory mutation at positions 79–90. Composition at sites with single mutations (positions 76–93) were: non-canonical pair G–U in V1-I, V1-IIA and V1-IIIA; canonical pair A–U in V1-IIB and V1-IIIB; and canonical pair G–C in V1-IIIC and V1-IV. The only non-canonical pair found in these structures was type G–U. A previous comparative study (Gutell et al., 1994) has shown that non-canonical pair G–U appears to play an incidental role in tRNA structure, but also occurs frequently in rRNAs. The so-called D-type (dominant) G–U pairs do vary in composition, interchanging with canonical pairs, in almost all groups across the spectrum of bacterial 16S rRNAs, but G–U occurs at significant levels and is the predominant composition in many of them. The composition at sites 76–93 found in the present study, is consistent with the observation of Gutell et al. (1994) for the sequences deposited in the databases. Stem V1-I, previously described in five rrn operons of strain PK3 (Cilia et al., 1996), was only found in strain ECOR42 (Table 2). The structure of stem V1-II was determined in all strains except ECOR35. V1-III (A, B and C) was only found in strains ECOR52, ECOR49 and ECOR35, respectively (Table 2). The sequence of V1-IV was frequently retrieved from strains ECOR35 and VECOR49, but was once also detected in ECOR56 (Table 2). It is worth noting that, in the study of Cilia et al. (1996), only two different sequences for this region were described.

Sequence diversity of stem V6

The two primary structures found in stem V6, of five primarily selected strains, showed nucleotide variation at nine sites (Fig. 2). Three mutations were compensatory: canonical pairs U–A and G–C of V6-I (positions 1010–1019 and 1002–1038) change to C–G and A–U in V6-II; non-canonical G–U (positions 1006–1023) also changes to C–G in the latter. Single mutations appear in three sites (positions 1020 to 1022): their compositions in V6-I were G, A and A, but V6-II showed A, U and U. Due to these changes, non-canonical pair U–G (positions 1009–1020) was replaced by U–A, but two pairs U–A (positions 1007–1022 and 1008–1021) were substituted by non-canonical U–U pairs. The sequence of stem V6-I was determined in all clones of strains ECOR52 and ECOR56; whereas all those from ECOR35 showed stem V6-II. Strains ECOR42 and ECOR49 showed clones of both types (Table 2). PCR using primer 16SOF and probe V6-I yielded an expected product of about 1000 bp when DNA template was from strains ECOR42, ECOR49, ECOR52 and ECOR56 but not
Table 2. Combination types of stems V1 and V6 determined in the cloned 16S rRNA genes from E. coli (ECOR) strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>No. of clones</th>
<th>Stem V1</th>
<th>Stem V6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td>III</td>
</tr>
<tr>
<td>ECOR 52</td>
<td>3</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECOR 56</td>
<td>5</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECOR 49</td>
<td>1</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECOR 42</td>
<td>2</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECOR 35</td>
<td>10</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECOR 24</td>
<td>9</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>PK3*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rRNA</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rnmB</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rnmC</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rnmE</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rnmD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rnmG</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rnmH</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Data obtained from Cilia et al. (1996) for the rnm operons of E. coli strain PK3.

from ECOR35; whereas probe V6-II gave the same product only for strains ECOR35, ECOR42 and ECOR49 (Fig. 3). Sequence of V6-II has previously been described in operon rnmH of strain PK3 (Cilia et al., 1996). A comparative analysis with all updated entries of the Ribosomal Database Project (RDP) (Maidak et al., 1994) showed that the sequence of V6-II is common in many other genera (viz. Xylella,
**Distribution of V6 structures in the ECOR strains**

A PCR test with V6-specific probes was applied to the 72 ECOR strains and distribution of these forms is shown in the MLEE tree (Herzer et al., 1990) of Fig. 3. Eleven strains showed only V6-II, 24 had V6-I and 36 revealed both types. All strains of MLEE group B2 showed only V6-I and a subcluster of six strains (ECOR35, ECOR36, ECOR38, ECOR39, ECOR40 and ECOR41) of group D only showed V6-II. It is remarkable that most strains (72%) from other MLEE groups contained rrr operons with both V6-I and V6-II. Strain ECOR24 did not yield PCR products with probes V6-I and V6-II; therefore, 12 cloned PCR-amplified 16S rRNA genes from this strain were subjected to sequencing of V1 and V6 variable regions.

Two clones exhibited a new V1 structure, V1-IIIC (Fig. 1), which differs from V1-IIA at a single site (position 93). In ECOR 24, stem V6 showed an as-yet-undetermined primary structure, V6-III (Fig. 2), which introduces three more variable sites with respect to V6-I and V6-II (positions 1002, 1011 and 1018). This loop showed the same pairs A-U and C-G (positions 1002–1038 and 1006–1023) as V6-II; the same pairs U-A and U-G (positions 1008–1021 and 1009–1020) as V6-I; but a unique composition in three canonical pairs exists: C-G, G-C and U-A (positions 1007–1022, 1010–1019 and 1011–1018). Comparative analysis of V6-III (22 bp; 1004–1025) with the RDP showed no nucleotide differences from those of *Shigella dysenteriae* and *Pasteurella haemolytica*, but only two differences from those of other genera (viz. *Actinobacillus*, *Haemophilus*). Specific V6-III probe indicated that five more strains (ECOR32, ECOR70, ECOR72 and ECOR42; Fig. 3) contain this sequence.

**Combinations of V1 and V6 structures**

Combinations of V1 and V6 structures found in the ECOR strains under investigation are in Table 2. All strains showed more than one combined structure (excluding subtypes A, B and C): eight clones from ECOR56 showed the combination V1-II/V6-I and one clone was V1-IV/V6-I; 10 clones from ECOR35 were V1-IV/V6-II and three clones showed the combined sequence V1-III/V6-II; three clones ECOR52 were V1-III/V6-I and three were V1-III/V6-II; three clones ECOR42 were V1-III/V6-II; two were V1-I/V6-II and one clone was V1-II/V6-I; nine clones from ECOR24 were V1-I/V6-III, and three were V1-II/V6-III. The most diverse combinations were retrieved from strain ECOR49: three clones were V1-III/V6-I, four were V1-IV/V6-I, two were V1-IV/V6-II, and one clone showed the combined sequence V1-II/V6-II. This last sequence was the only one found in the 56 clones that, unexpectedly, is identical to those of *Salmonella giv'ee* and *Salmonella* sp. (ATCC 9712).

**DISCUSSION**

Relatively high levels of 16S rRNA sequence heterogeneity between the different members of the *rrn* multigene family have already been described for several bacterial species (Ash et al., 1991; Dryden &

Thus, it seems not to be a rare phenomenon in bacteria. Sequencing of 16S rDNA for the seven rrn operons of E. coli strain PK3 showed that most nucleotide changes are not dispersed, but they were found to be grouped into two distinct domains of the molecule (Cilia et al., 1996). Almost complete 16S rDNA sequences of two clones of ECOR35 and ECOR52, and one clone of ECOR49 (not shown), confirm that two main variable regions exist and that they are involved in hairpin structures that seem to conserve base-pairing. A total of 56 clones from five ECOR strains were selected in a two-step analysis to determine sequence variation of these stems located at regions V1 and V6 of the 16S rDNA.

rRNA chronometers are considered to be ‘mosaics’ of sequences with very different degrees of variation, allowing most kinds of phylogenetic distance to be measured. Stems V6 and V1 are the more important stretches for intraspecific sequence variation within E. coli. Stem V6 showed three primary structures although there are 12 variable sites (Fig. 2). Eight different primary structures of stem V1 were determined in the selected clones from these six E. coli strains, in contrast with only two found in a previous study (Cilia et al., 1996). Stem V1 showed six sites of change but the eight sequences only differ from each other by a compensatory mutation of a canonical pair or by a single nucleotide at each site of a non-canonical pair (Fig. 1). Of course, if all these nucleotides are free to change in any direction following a random process, many more different V1 sequences should be found. Functional constrictions (the complexity of the translational apparatus) may not allow all possibilities and secondary structure must be conserved in these stems. In spite of many structure/function limitations, an interesting level of diversity in stem V1 actually exists and differences may result from single evolutionary events. In view of this diversity, it is more likely that all V1 stems in E. coli occurred spontaneously by mutation rather than by recombination as previously suggested for operons rrnD and rrnG in strain PK3 (Cilia et al., 1996).

The rrn operons evolve in concert by a homogenization process, presumably by molecular gene conversions (Dover, 1982, 1987; Ohta, 1991); otherwise the members of this multigene family would show a great level of sequence divergence, each recording the phylogenetic history of a single gene. Variability between members of an rrn family will depend on a relative proportion of rates: fixation of point mutation and homogenization. Stem V1 in E. coli seems to be subject to a rate of mutation relatively higher than that of homogenization which results in a considerable degree of sequence diversity. In stem V6 however, homogenization may have already worked to give, by majority, two functionally advantageous forms (V6-I and V6-II) between other possible structures initially originated by mutation in these isolates. Other V6 structures have existed in this species as five strains showed V6-III and several different sequences of this stem–loop have been found in other species. Obviously mutation does not occur equally in all operons; therefore the multigene family may conserve one or more copies with primitive sequences. Homogenization processes may go in different directions but, in the end, most or all rrn operons tend to show the same (perhaps identical) sequence. The finally resolved sequence in the rrn family has to be that of an available member (in other words, a single copy may be favoured by co-evolution of the translational apparatus) and the other rrn members be converted according to the sequence of such a copy. If the sequence of the ‘favoured’ copy has altered its composition, this derived character becomes fixed in the new lineage. However, if a primitive sequence was homogenized, the new generation retains the ancestral structure. As homogenization does not affect the entire length of the 16S rRNA equally (Cilia et al., 1996; and results from the present study), mosaics of evolved and ancestral variable regions may appear in some rrn operons of certain strains. Stems V6-I and V6-II have been retrieved from E. coli; strains ECOR52 and ECOR56 only show the sequence V6-I, ECOR35 only V6-II and ECOR42 and ECOR49 contain both types of V6 structures. Homogenization in strains ECOR42 and ECOR49 has not yet been completed and its tendency towards V6-I or V6-II is not clearly defined from the present data. The ‘favoured’ copy in strains ECOR52 and ECOR56 should be V6-I, and operons of ECOR35 are probably evolving in concert towards V6-II sequence.

The V6-specific test showed that half ECOR strains contained both V6-I and V6-II (Fig. 3). However, all strains of MLEE group B2 had only V6-I, indicating that this is a homogeneous group with a highly clonal structure, as previously described by sequencing of the 16S–23S intergenic spacer region (ISR) (Anton et al., 1998; Garcia-Martínez et al., 1996), RAPD and RFLP of total DNA (Desjardins et al., 1995), and electrophoretic patterns of carboxylesterase B (Picard et al., 1993). This uniformity was also demonstrated by six strains of a subcluster of MLEE group D, which had only V6-II, and this correlates with the finding of a unique sequence in the ISR limited to only these six strains (Garcia-Martínez et al., 1996). Most, if not all, rrn copies in ECOR24 contained V6-III as this sequence was retrieved from the 12 clones and probes V6-I and V6-II did not yield PCR products. Sequence V6-III was also encountered in other genera (data from RDP), even in distantly related taxa such as Actinobacillus (although two nucleotide differences were shown). This character homogenized in ECOR24 was not frequent in the ECOR collection (five strains; Fig. 3): strains ECOR42, ECOR72 and ECOR32 simultaneously exhibited V6-I, V6-II and V6-III; strain ECOR70 had V6-II and V6-III, but none showed V6-
III and V6-I. It is noteworthy that V6-III was not determined in ECOR42 by sequencing. Further V6 structures different to those determined in the present study possibly exist in *E. coli*.

Comparison of V6-I sequence with data of the RDP showed that complete similarity is only found with sequences of *E. coli*, indicating that this derived character originated in the evolution of this species. The stem V6-II is a more general character, the primitive, as it was also found in many species of other genera. Therefore, ribosomal operons of *E. coli* with V6-II have nucleotide composition of variable positions in common with these organisms and this is why some clones obtained from *E. coli* may be more related to the rRNA of a different species (and genera) than to other members of the *rrn* family. The phylogenetic effect produced by homogenized primitive sequences may be similar to that of homoplasies (convergences, parallelisms and reversals; Stewart, 1993); such a thing is highly unlikely by spontaneous mutation. Evidence for these distorting effects is a loss of ultrametricity in phylogenetic inference. A poorly ultrametric phylogenetic tree usually shows very uneven tips (Sneath, 1993) and this was found in the analysis of phylogenetic relationships among the seven operons of *E. coli* PK3 (Cilia et al., 1996); sequences of some *rrn* have affinity for those of *Shigella* spp. and, moreover, *rrnH* of strain PK3 (having V6-II) is closer to a different species than to other *rrn* copies from the same multigene family. A clone of ECOR 49 illustrates well this kind of chronometric distortion. The clone showed the sequence combination V1-IIA/V6-II, the same as that found in *Salmonella give* and *Salmonella* sp. (ATCC 9712), therefore this clone will show a relatively close relationship with these two *Salmonella* species in an rRNA-based phylogenetic tree. PCR using the three V6 probes has confirmed that ECOR 35 only contains the sequence of V6-II. As homogenization may take different directions for each domain, this strain will have evolved parts of the molecule combined with the more primitive sequence V6-II; the latter would make the distortion effect of a homoplasy, responsible for an uneven tip for ECOR 35 in the phylogenetic tree. Therefore, if unequally directed homogenization processes are involved in concerted evolution, the homoplasy effect may be more likely than when only spontaneous convergences or transference of genetic material are considered.

Long-term evolutionary analyses (above the species level) include sequences that have had an enormous chance of repeated mutations in the same variable positions, and distortions may have been corrected. As the size of the 16S rRNA is large, many independent domains exist and numbers of chronometric changes between distantly related microorganisms are high, recent distortions due to homogenization of short primitive sequences do not alter the mean phylogeny, and trees are ultrametric. In short-term evolution, these anomalies are more evident: the number of nucleotide changes is too low. In addition, chronometric distortions due to homogenized primitive sequences may therefore be more likely for large multigene families, and not possible for bacterial organisms with only a single *rrn* copy. All these considerations may explain some phylogenetic rate inconsistencies found between closely related species with multiple *rrn* copies and their consequent poorly ultrametric phylogenetic trees (Ash et al., 1991; Fox et al., 1992; Martínez-Murcia, 1993; Martínez-Murcia et al., 1992; Rössler et al., 1991; Ruimy et al., 1994; Sneath, 1993). Nevertheless, intraspecific sequence variation does not affect the species delineation if a number of consistent species-specific nucleotides exist.

rRNAs are so far the most informative of all known genetic sequences for measuring phylogenetic distances. Size and domain composition, universal distribution, age and functional consistency of the ribosome, conserved secondary structure and the sequence not subjected to degenerated code, are important advantages over other molecular chronometers. Reconstruction of long-term phylogenies seems a reasonably acceptable hypothesis of the bacterial history as they have been supported by similar tree topologies derived from other different and duplicated genes (Cammarano et al., 1992; Iwabe et al., 1989). At present most studies on bacterial systematics include sequencing data, as a polyphasic approach to assess diversity has been recommended (Muray et al., 1990). With the development of rapid methods for rRNA (or gene) sequencing (Böttger, 1989; Lane et al., 1985), a high number of 16S rRNA sequences have been described and deposited in databases over the last few years. However, analyses at the intraspecies level have received little attention and, probably more importantly, the problem derived from the organization of rRNA as a multigene family subjected to homogenization, has been completely ignored in most sequencing analyses. Different authors have interpreted controversial data suggesting an interspecies recombination of rRNA genes in the ancestor of a genus (Sneath, 1993), or between different genera (Cilia et al., 1996), and considered that horizontal gene flow provides a unique explanation for a number of contradictory clocks (Syvanen, 1987). Although transference of genetic material is possible, and has to be taken into consideration as an alternative explanation, to our knowledge there is not enough evidence to suggest that this process has randomized the origins of genes (Olsen & Woese, 1993), and transference of an *rrn* copy may be a rare phenomenon. Therefore the widespread occurrence of stem V6-II in many different taxa becomes very difficult to explain by this mechanism. Analysis in the present work has shown data to support that, in *E. coli*, rRNA sequence diversity at some stems–loops (viz. V1) comes from random mutation and that different directions of partial (i.e. V6) homogenization of the *rrn* multigene family may cause some kind of chronometric distortions and pictures with low ultrametricity. Further investigations of *rrn* families in a wide range of very
closely related microbial taxa are desirable for a deeper understanding of molecular clock behaviour and evolution of bacterial species.

ACKNOWLEDGEMENTS

This work was supported by grant PM95-0111 of the CICYT. Secretarial assistance from K. Hernández is gratefully acknowledged.

REFERENCES


