

Intraspecific Diversity of the 23S rRNA Gene and the Spacer Region Downstream in *Escherichia coli*

ANA I. ANTÓN, ANTONIO J. MARTÍNEZ-MURCIA,[†] AND FRANCISCO RODRÍGUEZ-VALERA*

Unidad de Microbiología, Centro de Biología Molecular y Celular, Universidad Miguel Hernández, Campus de San Juan, 03550 San Juan, Alicante, Spain

Received 28 December 1998/Accepted 3 March 1999

The molecular microevolution of the 23S rRNA gene (*rrl*) plus the spacer downstream has been studied by sequencing of different operons from some representative strains of the *Escherichia coli* ECOR collection. The *rrl* gene was fully sequenced in six strains showing a total of 67 polymorphic sites, a level of variation per nucleotide similar to that found for the 16S rRNA gene (*rrs*) in a previous study. The size of the gene was highly conserved (2902 to 2905 nucleotides). Most polymorphic sites were clustered in five secondary-structure helices. Those regions in a large number of operons were sequenced, and several variations were found. Sequences of the same helix from two different strains were often widely divergent, and no intermediate forms existed. Intercistronic variability was detected, although it seemed to be lower than for the *rrs* gene. The presence of two characteristic sequences was determined by PCR analysis throughout all of the strains of the ECOR collection, and some correlations with the multilocus enzyme electrophoresis clusters were detected. The mode of variation of the *rrl* gene seems to be quite similar to that of the *rrs* gene. Homogenization of the gene families and transfer of sequences from different clonal lines could explain this pattern of variation detected; perhaps these factors are more relevant to evolution than single mutation. The spacer region between the 23S and 5S rRNA genes exhibited a highly polymorphic region, particularly at the 3' end.

The rRNA genes are the most frequently used markers for the study of molecular phylogeny for reasons that have been repeatedly reviewed (1, 30, 32), although their applicability to establish organismal phylogenies in prokaryotes has been recently contested (19, 26, 33). The *rrn* operons are among the few genes in prokaryotic genomes that are often repeated. In *Escherichia coli* seven *rrn* operons are located throughout the half of the chromosome containing the origin of replication (6). Little is known about the mode of evolution of these genes. A standard way to approach this problem is to study sequence variation within populations of relatively recent divergence, so that the general pattern of molecular evolution can still be discerned (31). Lately, we have studied the variation affecting different regions of the rRNA operons in *E. coli* and related species (2, 10, 22, 23), specifically, the *rrs* gene and the internal transcribed spacer (ITS) region between the *rrs* and *rrl* genes. As expected, relatively little variation of single nucleotide substitutions for the *rrs* gene was found, and higher variation for the *rrs-rrl* ITS was found. There were hypervariable regions in both elements where clustered nucleotide substitutions occurred, generally conserving the rRNA secondary structure (compensatory mutations). In some regions, such as helix V1 of the *rrs* gene, several variants were found. The different sequence variants differed by single or compensatory mutations, and possible relationships among the different sequences were shown (23). This mode of gradual change fits well with a model of evolution by neutral change and genetic drift. However, the most frequent situation in both the *rrs* gene and the *rrs-rrl* ITS (2) was the finding of a restricted number of highly

divergent sequences that maintained the same secondary structure. This “jumpy” mode of variation is difficult to explain by neutral-change models. Sometimes the sequences found are similar to those of other species (5), and some researchers have suggested lateral transfer and recombination to explain their presence in *E. coli*. Regardless of their origin, molecular coevolution with ribosomal proteins or the transcript processing machinery could be an important reason for the preservation of these sequences (8).

The most remarkable difference found between the *rrs* gene and the *rrs-rrl* ITS was the presence in the latter of relatively large insertions and deletions, as exemplified by the *rsI* loop, that alter the secondary structure dramatically. A mode of evolution in which cumulative insertions and deletions would eventually alter the overall sequence of the *rrs-rrl* ITS was suggested (2, 22). This would explain the great variation found when organisms with increasing phylogenetic distances are compared, as opposed to the *rrs* gene, in which variation on a larger phylogenetic scale is relatively proportionate.

Here, we have studied another region within rRNA operons that includes the *rrl* gene and the spacer region downstream (*rrl-rrf* ITS). The *rrl* gene is apparently less restricted in size variation (11), and a wide divergence exist throughout the class *Proteobacteria* (21). In addition, the *rrl* gene is also assumed to be more variable at the sequence level than the *rrs* gene, particularly for shorter phylogenetic distances. Representatives from different clusters determined by multilocus enzyme electrophoresis (MLEE) of the *E. coli* ECOR collection (25), plus one O157 serotype isolate, were selected for sequencing. Hypervariable regions have been sequenced to assess intercistronic heterogeneity. Finally, the determined variants of two helices were examined by PCR analysis throughout the ECOR collection.

MATERIALS AND METHODS

Bacterial strains and cultivation. *E. coli* strains of the ECOR collection, comprising isolates from a wide variety of hosts and geographic regions (25),

* Corresponding author. Mailing address: Unidad de Microbiología, Centro de Biología Molecular y Celular, Universidad Miguel Hernández, Campus de San Juan, Apartado 18, 03550 San Juan, Alicante, Spain. Phone: 965919451. Fax: 965919457. E-mail: FRVALERA@UMH.ES.

[†] Present address: Departamento de Biotecnología, Universidad de Alicante, E-03080 Alicante, Spain.

	Helix 25	Helix 45	Helix 63	Helix 79	Helix 98
	11223333555555555555	89111111111111111111	111111111111111111	1112222222222222	2222222222222222
	83468155514444444455	4377771122301482223333	67783600011124444	999900372456267712345678902	6614581909062826781456502323359135513679457903
<i>E. coli</i> K12 <i>rrlA</i>	ATACGTAGCCACGCTTAGGCTTAA	ATGCCTATCACCGATTTCGTAGGTA	TAGATC	GAGTAG	CTCTGA
<i>E. coli</i> K12 <i>rrlB</i>C.....	GT.CC.....	TCCCGGA.CA.....	CT.AG
<i>E. coli</i> K12 <i>rrlC</i>AC.AT.....	GT.CC.....	CCA...G..G..A.....
<i>E. coli</i> K12 <i>rrlD</i>C.....	GT.CC.....	CA.A.....	CAGTT	CT.AG
<i>E. coli</i> K12 <i>rrlE</i>C.....	GT.CC.....	TCCCGGA.CA.....
<i>E. coli</i> K12 <i>rrlG</i>C.....	GT.CC.....	TCCCGGA.CA.....	CT.AG
<i>E. coli</i> K12 <i>rrlH</i>	...G.C....T.....	C.GT.C.A.....	CA.....
ECOR 10 <i>rrlD</i>C.....TGC.....	ATCCCGGA.CAAA.....	CT.AG
ECOR 24 <i>rrlB</i>	.AT..C.AT.CTCT..TATGG..	G.GC.CTGC..A..TCC.GGACCA.A..
ECOR 35 <i>rrlC</i>	G...C...TCTCT...ATGG..	GT.C.CTG.....	...C..G.CCG.G.....	GATA.T
ECOR 40 <i>rrlB</i>	.ATG...A..CTCT.A-ATGG..	...TA.....	TCCCGGA.TA..G.CCGCG.....
ECOR 52 <i>rrlD</i>	G..G..GAT.CTCT.A-ATGG..GTC.G.....	...C...G.....
A8190 <i>rrlC</i>	.AT..C....CAGAA.TCT..	..-G..TGC.....	TCCCGGA.CCA...G..G..	CT.A

FIG. 1. Variable positions in the *rrl* genes of *E. coli* K-12 and other strains used in this study. Periods indicate the same nucleotide composition as the *rrlA* sequence of *E. coli* K-12 (*Escherichia coli* Database Collection). The numbers at the top indicate positions in the sequences reported by Brosius et al. (3). Insertion of nucleotides G and C between positions 545 and 546 is indicated by the arrow.

were used. Strain A8190 of *E. coli* serotype O157:H7 (kindly provided by T. S. Whittam, Pennsylvania State University) was also included. Strains were grown in Luria-Bertani (LB) medium for 18 h at 37°C and stored at -80°C in 15% glycerol.

DNA extraction. Bacterial genomic DNA was extracted by using the InstaGene DNA Purification Matrix (Bio-Rad Laboratories, Inc.) in accordance with the manufacturer's indications. DNA concentrations were spectrophotometrically estimated and diluted to 100 ng/μl.

PCR amplification. Long-distance PCR was used to amplify the complete *rm* operon and the upstream flanking sequence of ca. 1,000 bp. Previously described primers (2) designed on the basis of specific sequences located upstream of the *rrs* gene of *E. coli* K-12 were combined with primer 5SR (5'-ATGCCTGGCAG TTCCTACT-3'). Reaction mixtures of 50 μl contained 1 μl of genomic DNA, 60 mM Tris-SO₄ (pH 9.1), 18 mM (NH₄)₂SO₄, 1.8 mM MgSO₄, 0.2 mM each deoxynucleotide, 0.2 μM each primer, and 2 U of Elongase Enzyme Mix (Gibco BRL, Life Technologies Ltd.). PCR was performed with a PTC-100 thermal cycler (MJ Research, Inc.). The mixtures were subjected to 35 cycles of 94°C for 30 s, 55°C for 30 s, and 68°C for 6 min. A final extension step of 68°C for 5 min was allowed. Probes designed on the basis of two variable regions of the *rrl* gene were H25R I (5'-CACACGCCTAAGCGTGCTCC-3'), H25R II (5'-CACCCC ATTAAGAGGCTCCC-3'), and H25R III (5'-GTACACAGATTGCTCTGT G-3') for helix 25 and H63R I (5'-CAGTCCATCCGCGAGGGAC-3') and H63R II (5'-CAGTCCACGACGAAGTCCG-3') for helix 63. The PCR conditions used with these probes were as previously described (2). PCR products were analyzed in 1% agarose gels, stained with ethidium bromide, and visualized on a UV transilluminator.

Sequencing of the *rrl* gene and the *rrl-rrf* ITS. PCR products were purified by using a QIquick PCR Purification Kit (Quiagen) in accordance with the manufacturer's protocol. Nucleotide sequences were determined in an ABI PRISM 377 sequencer (Perkin-Elmer) as recommended by the manufacturer. The sequencing primers for the *rrl* gene were 23SOF (5'-GTGAGTCTCTCAAATTT TCG-3'), 23SF (5'-AGTAGCGGCGAGCGAACGGG-3'), 23S6F (5'-GCGT ACCTTTGTATAATGG-3'), 23S8F (5'-ATAGCTGTTCTCCCGAAA3'), 23S11F (5'-AAGAAAGCGTAATAGCTCAC-3'), 23S16F (5'-TACCCCAA ACCGACACAGG-3'), 23S20F (5'-TAGCGAAATTCCTTGTCCGG-3'), 23S23F (5'-GGGTAGTTGACTGGGGCGG-3'), and 23S26F (5'-CAAGGGTATGGC GTTTCGCC-3'). Primer 5SR was used to sequence the *rrl-rrf* ITS.

Data analysis. Alignment and phylogenetic analysis of the *rrl* gene sequences were carried out by using the programs ClustalW and Mega (Molecular Evolutionary Genetics Analysis, Version 1.01), respectively. Comparisons with sequences in the GenBank database were achieved by BLAST searches (BLASTN). The GenBank accession numbers of the *rrl* genes of strains ECOR 35, ECOR 52, ECOR 40, ECOR 10, ECOR 24, and A8190 are AF053963 to AF053968, respectively.

RESULTS

The six *E. coli* strains were subjected to locus-specific PCR using primers located upstream from each *rm* operon. The strains were selected considering the diversity found in both

the *rrs* gene and the *rrs-rrl* ITS (2, 23), as well as their relationships shown by MLEE (16). ECOR 10 is highly related to *E. coli* K-12 and was used as a reference. ECOR 24 was selected because it contains unusual intergenic homogeneity and the *rrl* loop in all of the *rrs-rrl* ITSs. ECOR 35 is relatively homogeneous regarding differences among the *rm* operons at the level of the *rrs-rrl* ITS. ECOR 40 belongs to an MLEE cluster with a peculiar insertion in the *rrs-rrl* ITS that is also found in *Salmonella* spp. and contains a high level of intergenic heterogeneity. ECOR 52 is a representative of the MLEE B2 group that represents a highly homogeneous phylogenetic cluster that lacks the variants found in other MLEE clusters. Finally, A8190 is a representative of verotoxigenic *E. coli* serotype O157, which has recently been shown to be widely divergent from other *E. coli* lineages (27). PCR amplification yielded amplicons of ca. 6 kb. Complete *rrl* gene sequences of a single operon from each strain were determined in order to examine representatives of the operons containing one (*rmC* and *rmB*) and two (*rmD*) tRNA genes. Figure 1 summarizes the 67 polymorphic sites found from alignment of the six *rrl* genes fully sequenced. The pattern of variation was highly reminiscent of that found for the *rrs* gene (23), and the percentage of polymorphic sites is about the same. All of the *rrl* genes were almost identical in size, 2902 to 2905 bp. Intervening sequences, like those found in *Salmonella rrl* genes (4), were not detected. They were not found in the *rrs* genes of *E. coli* strains (5, 23) either, and if they are present at all in *E. coli*, they should be much rarer than in *Salmonella* spp.

Most polymorphic sites clustered in five secondary-structure loops, viz., helices 25, 45, 63, 79, and 98 (helix numbering is that of Larsen [20]; Fig. 1 and 2). In helices 25, 63, and 98, although the secondary structure is well preserved, the sequence is often remarkably divergent. A moderate degree of polymorphism was observed in helices 45 and 79, and some intermediate forms could be related by single and compensatory mutational events. In helix 25, the three forms detected had a distribution clearly associated with the MLEE clusters. Partial sequences of the most variable regions were obtained from a larger number of *rm* operons of the six strains previously amplified by PCR using operon-specific primers. The types of primary structures detected are summarized in Table

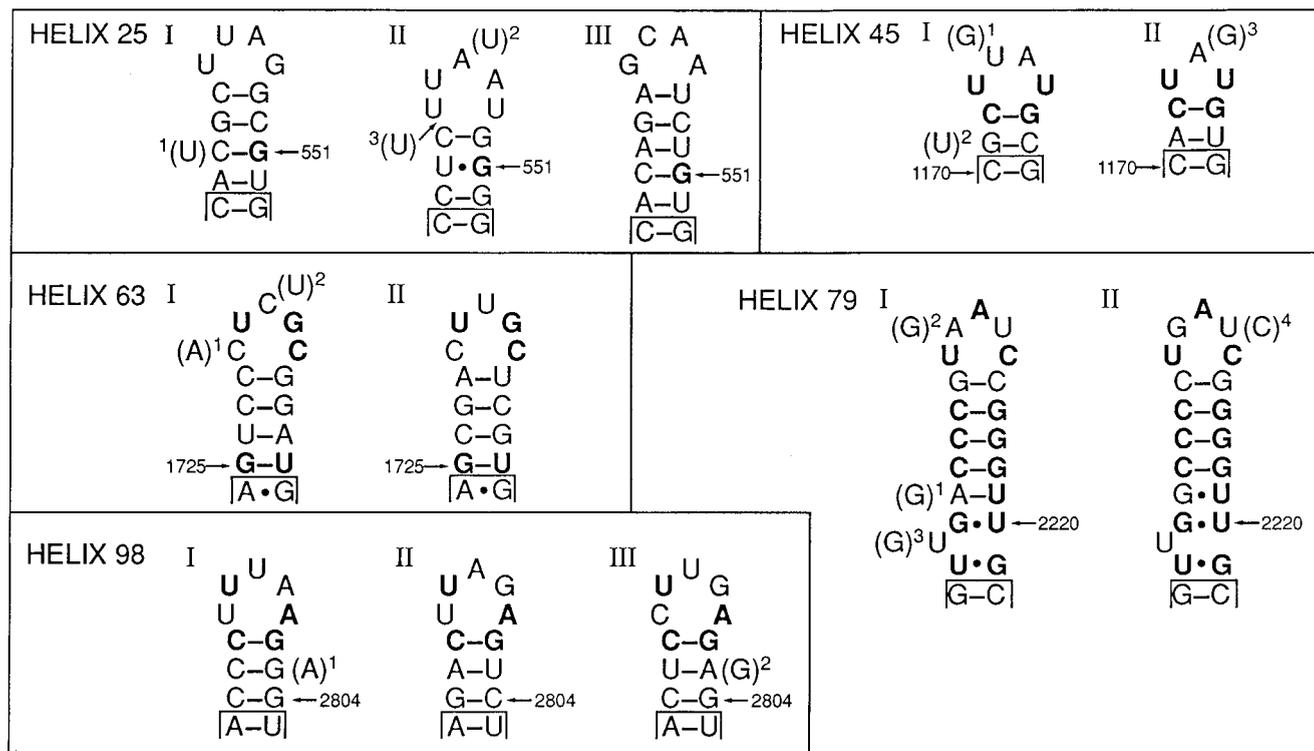


FIG. 2. Secondary structures of variable regions in the *rfl* genes from *E. coli* strains. The numbers indicate positions reported by Brosius et al. (3). Conserved nucleotides are in boldface. In parentheses are the nucleotides that have been determined in the *rfl* genes of the following strains: helix 25, (U)¹ in *rflH* of ECOR 40, (U)² in *rflB*, *rflC*, *rflE*, and *rflG* of ECOR 35, and a (U)³ insertion in *rflB*, *rflC*, and *rflD* of ECOR 24 and *rflG* of A8190; helix 45, (G)¹ in *rflB*, *rflG*, and *rflH* of ECOR 10 and *rflB*, *rflC*, and *rflD* of ECOR 24, (U)² in *rflC* of ECOR 24; and (G)³ in *rflA* and *rflC* of A8190; helix 63, (A)¹ in *rflC* of ECOR 40 and (U)² in *rflB* and *rflC* of ECOR 10 and *rflB*, *rflC*, and *rflD* of ECOR 24; helix 79, (G)¹ in *rflA*, *rflC*, *rflD*, and *rflE* of ECOR 52 and (G)² and (G)³ in *rflB* and *rflC* of ECOR 10 and *rflA*, *rflC*, and *rflG* of A8190; helix 98, (A)¹ in *rflC* of A8190 and (G)² in *rflA* and *rflE* of ECOR 52.

1. These sequencing results confirmed the polymorphism already found in the fully sequenced *rfl* genes; no new forms were detected in the hypervariable helices under study (Fig. 2 and Table 1). Except for helices H25 II, H25 III, and H98 II, all of the structures were previously found in *E. coli* K-12 (data from the *Escherichia coli* Database Collection). A remarkable aspect of the results presented in Table 1 is the high degree of interclonal homogeneity found within each strain. This homogeneity has been confirmed by a PCR test using specific probes for helix 25 and, to a lesser degree, helix 63 (Fig. 3). *E. coli* K-12 and strain ECOR 10 (which are highly related) were the most heterogeneous, in agreement with previous analysis of the *rrs* gene (23) and the *rrs-rfl* ITS (2).

For helices showing widely divergent sequences (helices 25 and 63), PCR primers were designed and used as probes to screen the ECOR collection in order to evaluate their distribution throughout the phylogenetic expanse of the species (Fig. 3). The sequence of helix 25 I, found in *E. coli* K-12, was restricted to MLEE cluster A plus two strains of B1. The probe for ECOR 25 II hybridized with most of the strains of groups B2, D, and E. Altogether, the distribution of these different forms seemed to correlate quite well with the clusters defined by MLEE. Finally, the probe for the sequence of helix 25 III, found originally in strain A8190, was restricted to a group of strains within cluster B1 of MLEE and only two strains outside this cluster, ECOR 12 of cluster A and ECOR 37 of cluster E. All of the other representatives of group E were negative with this probe. In a previous study of *mdh* sequences using representatives of *E. coli*, only ECOR 37 fell within the cluster that

includes representatives of the O157:H7 serotype (27). Only four strains in the ECOR collection were positive by more than one H25 probe, showing remarkable interclonal homogeneity at this level. A different result was obtained with probes for the two sequences found in helix 63. Twenty-five strains were positive with both probes, 25 were positive only by the probe for helix 63 I (half of them belonging to group B2), and the remaining 22 strains were positive only by the helix 63 II probe. In this case, interclonal heterogeneity must be as high as that of *E. coli* K-12 (Table 1).

The sequences found in the hypervariable loops were compared to those in databases to assess similarity to the equivalent loops in other bacterial species. No significant match was found, with the single exception of helix 98 II, which was most similar (only two nucleotides were different) to the equivalent helix of *Klebsiella pneumoniae*. However, the tetranucleotide (AACG) at the end of helix 25 III was found in many members of the class *Proteobacteria*, including *Thiobacillus cuprinus*, *Nannocystis exedens*, and *Stigmatella aurantica*.

23S-5S rRNA gene ITS region (*rfl-rrf* ITS). This spacer region is highly conserved among the *rrn* operons of *E. coli* K-12. In this strain, 57 of the 94 nucleotides of this ITS are presumed to form a secondary structure by pairing with the corresponding stretch downstream of the 5S rRNA gene (*rrf*) (3). However, strains in the present study showed significant variations affecting mostly the region that does not contribute to the secondary structure of the transcript (Fig. 4A). Particularly, one of the operons of ECOR 40 (group D) and the three operons of ECOR 52 (group B2) are widely divergent

TABLE 1. Determination of variable regions in the *rnl* genes of the ECOR collection strains used in this study and *E. coli* K-12

Strain and locus	Sequence present at ^a :											
	Helix 25			Helix 45		Helix 63		Helix 79		Helix 98		
	I	II	III	I	II	I	II	I	II	I	II	III
K-12												
<i>rnlA</i>	+				+			+	+			+
<i>rnlB</i>	+			+		+		+	+			+
<i>rnlC</i>	+			+		+		+	+			+
<i>rnlD</i>	+			+		+		+	+			+
<i>rnlE</i>	+			+		+		+	+			+
<i>rnlG</i>	+			+		+		+	+			+
<i>rnlH</i>	+			+		+		+	+			+
ECOR 10												
<i>rnlB</i>	+			+		+		+	+			+
<i>rnlC</i>	+			+		+		+	+			+
<i>rnlD</i>	+			+		+		+	+			+
<i>rnlG</i>	+			+		+		+	+			+
<i>rnlH</i>	+			+		+		+	+			+
ECOR 24												
<i>rnlB</i>	+			+		+		+	+			+
<i>rnlC</i>	+			+		+		+	+			+
<i>rnlD</i>	+			+		+		+	+			+
ECOR 35												
<i>rnlB</i>		+		+		+		+	+			+
<i>rnlC</i>		+		+		+		+	+			+
<i>rnlE</i>		+		+		+		+	+			+
<i>rnlG</i>		+		+		+		+	+			+
ECOR 40												
<i>rnlB</i>		+		+	+			+	+			+
<i>rnlC</i>		+		+	+			+	+			+
ECOR 52												
<i>rnlA</i>		+		+		+	+					+
<i>rnlC</i>		+		+		+	+					+
<i>rnlD</i>		+		+		+	+					+
<i>rnlE</i>		+		+		+	+					+
A8190												
<i>rnlA</i>			+		+	+		+	+			+
<i>rnlC</i>			+		+	+		+	+			+
<i>rnlE</i>			+		+	+		+	+			+
<i>rnlG</i>		+			+	+		+	+			+

^a A plus sign indicates the sequence found at the locus and helix in question.

from *E. coli* K-12 in their *rnl-rnf* ITS sequences. A remarkable interclonal heterogeneity was found affecting the determined *rnl-rnf* ITS of ECOR 40.

DISCUSSION

23S rRNA gene. Overall, the sequence variation found in the *rnl* gene was smaller than expected. The 23S rRNA gene is larger and shows a higher rate of variation than does the rRNA of the small subunit (11, 18). When different genera of the class *Proteobacteria* are compared, the similarity of the 16S rRNA genes is about 4% higher than that of the 23S rRNA genes (21). In a previous study (23), five complete *rns* gene sequences from three strains in the ECOR collection showed 34 polymorphic sites (unpublished data). In the present study, 67 polymorphic sites were detected in the six fully sequenced *rnl* genes. Therefore, at this level, variation per nucleotide seems to be

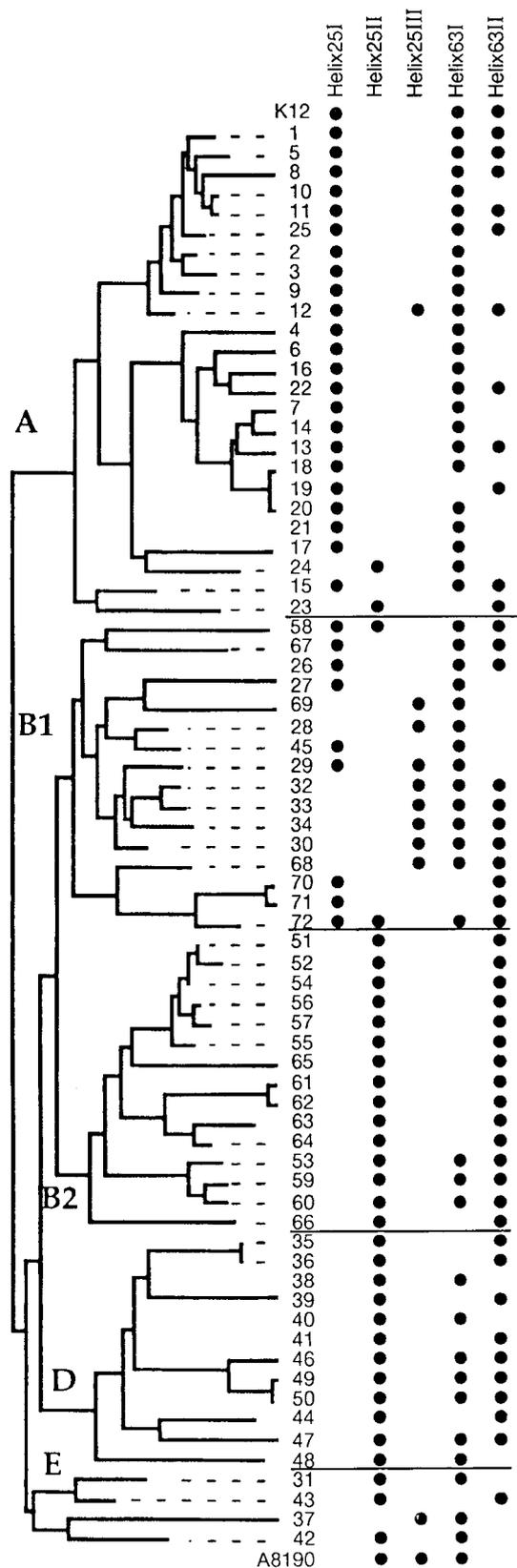


FIG. 3. Distribution of structures I, II, and III of helix 25 and structures I and II of helix 63 among the ECOR strains as revealed by a PCR test using sequence-specific probes. The MLEE relationships shown in the tree are those described by Herzer et al. (16) modified by the addition of *E. coli* K-12 and A8190.

A

```

E.coli K12 rrnA ACAACGCCGAAGCTGTTTTGGCGGATTGAGAGAA-GATTT-TCAGCCTGATA-CA-GATTAAATCAGAACG--CAGAAGCGGTCTGAT----AAACAGAATT
ECOR 10 rrnC .....A.....
ECOR 24 rrnC .....A.....
ECOR 35 rrnC .....A.....-A...C.TC...AAG.A...T...A...T.....T.....
ECOR 40 rrnC .....G.....-.....C.TC...CAAG.A...T.....GAC.G.T-...CA..A.GTG.TGAT.....
ECOR 24 rrnD .....A.....
ECOR 40 rrnB .....A.....
ECOR 52 rrnA .....A.....
ECOR 52 rrnE .....A.....
ECOR 52 rrnG .....T.....
A8190 rrnA .....G.....

```

B

```

E.coli K12 rrnA ACAACGCCGAAGGTGTTTTGGCGGATT-GAGAGAAGATTTTCAGCC-TGATACAGATTAATCAGAACG-CAGAAGCGGTC-----TGATAAAACAGAATT
ECOR 40 rrnC .....GA...C.TC-...CAAGT.AG.T.....GAC.G.T....CA..ACGT--GT.....
S.typhi .....A.....GA...C-...AC...T.C.GA.TTTCGG...AA...TAAAGATTCGTT.CCGGC.....

```

FIG. 4. Alignment of *rnl-rf* ITS sequences determined for ECOR strains and A8190 and for *rrnA* of *E. coli* K-12. (B) Alignment of *rnl-rf* ITSs of *E. coli* K-12, ECOR 40, and *S. typhi*. Mosaic similarities are underlined. Periods indicate the same nucleotide composition.

equivalent for both genes. Similarly, an unexpected conservation was found for the size of the *rnl* gene. Even disregarding the contribution of frequent intervening-sequence insertions, the size of the 23S rRNA molecule in members of the class *Proteobacteria* varies in a range of ca. 200 nucleotides (21). The size of the *rnl* genes within the ECOR collection varied by only three nucleotides; apparently, as happens with the spacer regions (2, 22), variability at the interspecific level cannot be used to infer intraspecific variability.

On the other hand, the hypervariable regions found correspond to the variable loops detected by sequence comparison of *rnl* genes from widely divergent *Proteobacteria*. Helices 25, 63, and 98, which contained most of the variable sites, are also the most variable helices at the level of different proteobacterial subclasses. This could be interpreted as just a reflection of the low functional restriction of these regions of the secondary structure. However, the pattern of variation found in these helices does not fit well with a model of neutral variation by genetic drift (14). In the latter, a more gradual variation should be expected (15). Instead, we found alternative sequences that are not similar but are apparently well preserved among strains. Some of the sequences found in these helices are not completely unique to *E. coli*. For example, the nonpaired nucleotide stretch at the end of helix 25 III was identical to that of distantly related genera of *Proteobacteria*. However, the small number of identical nucleotides seems to indicate that this reflects common ancestry and functional conservation rather than horizontal transfer. In helix 98 II, a nearly identical sequence for the whole loop is found in *K. pneumoniae*. In this case, horizontal transfer seems more likely given the high similarity over a wide stretch of nucleotides and the close relationship between the two species. Still, other regions seem to vary according to a gradual pattern. The sequence diversity found in helices 45 and 79 could easily have originated from single and compensatory mutations affecting these areas, and the same applies to the isolated nucleotide substitutions. In all of these aspects, the *rnl* gene is reminiscent of the *rs* gene, where hypervariable regions of both types have been found (23).

In the *rs* gene, intergenic heterogeneity in two variable regions was rampant, and most strains contained more than one of the sequence types detected (23). Something similar happens in helix 63 of the *rnl* gene. However, intergenic heterogeneity seems to be rather restricted in helix 25. As detected by sequencing and probing by PCR, most strains, including *E. coli* K-12 and ECOR 40, seem to have only one of the sequence types found, although both strains are particularly heterogeneous in other regions of the operon (2, 23). The distribution of sequence types for helix 25 follows the MLEE phylogeny quite well. For instance, helix 25 I seems characteristic of group A, and strains of clusters B2 and D only showed helix 25 II. The clonality of cluster B2 was also evidenced by a previous analysis of *rs* sequence diversity (23).

23S-5S rRNA gene ITS region (*rnl-rf* ITS). Sequencing of the *rnl-rf* ITS has revealed the existence of an undescribed hypervariable region affecting ca. 40 nucleotides before the 5' end of the *rf* gene (Fig. 4A). This region seems to be the most variable found throughout the *rrn* operon. We have found again a strong heterogeneity affecting ECOR 40, a strain that is highly heterogeneous in the *rs* gene and the *rs-rnl* ITS (2, 23). The mosaic relationships found in the spacer regions of *E. coli* K-12, ECOR 40, and *S. enterica* serovar *typhi* are remarkable (Fig. 4B). This situation is often found in comparisons of homologous genes of bacteria (9, 24, 28), and it has generally been attributed to recombination. However, recombination acting over such short stretches of nucleotides appears unlikely. In spite of the wide variation at the nucleotide sequence level, the size of the spacer was remarkably conserved for the strains sequenced. This is in contrast to the *rs-rnl* ITS, where size varies widely, even among orthologous (containing the same tRNA gene) operons (3).

Molecular evolution of the *rrn* operons in *E. coli*. Only restricted information regarding the sequence diversity in roughly half of the ribosomal operons studied here is available, because only the hypervariable regions were sequenced in more than one operon per strain. However, from the partial sequences and the PCR probe screening, it is quite apparent that this section of the ribosomal operon behaves similarly to

the other half. The information accumulated regarding the sequence variation in different strains of the ECOR collection allows us to draw a general picture of how the evolution of the operon has proceeded. As expected, variation is concentrated in specific regions, and the sequence variation preserves the secondary structure (i.e., covariation) (13). This much can be predicted from what is known about the variation at the interspecific level. However, there are two markedly different modes of variation. One is exemplified by helix V1 of the *rrs* genes or helices 79 and 45 of the *rnl* genes that exhibit a gradual variation that could be explained by a lack of functional restriction and genetic drift. Other regions, such as V6 of *rrs* genes or helices 25, 63, and 98 of *rnl* genes only showed a few different, widely divergent sequence types, but no gradual variation was present. In the latter, a high level of similarity to the equivalent region in other bacterial species is sometimes found. These widely divergent variations could have originated by mutation or by horizontal transfer. In any case, the lack of gradual variation among the ECOR representatives studied hints of a more restricted freedom of variation by functional restriction and/or coevolution with ribosomal proteins. The spacer regions vary in a similar pattern, with relatively few isolated nucleotide substitutions and most of the polymorphisms clustered in areas that sometimes also correspond to conserved secondary-structure loops. The *rrs-rnl* ITS contains important insertion and deletion events that are not found in the structural genes and could be the main force behind the wide divergence found when spacers from distantly related species are compared.

Another aspect of the evolution of the *rm* operons is inter-cistronic heterogeneity. The concerted evolution of the rRNA gene family has been shown to be essential in eukaryotes for the relatively slow rate of change of those genes and noncoding sequences (18). The resulting effect of molecular drive has been suggested to be an important force in evolution in general and in the maintenance of species coherence and even identity, i.e., low (or negligible) levels of intraspecific variation versus notable interspecific variation (7, 8). Based on the available preliminary analysis, inter-cistronic and intraspecific variation of the rRNA genes appears to be very small in those organisms (17). On the other hand, a more limited number of copies of the repeating units is present in prokaryotes (1 to 13 *rm* operons) (12) than in eukaryotes (typically in the range of hundreds). In addition, the panmictic structure found for most diploid multicellular eukaryotic populations cannot be extrapolated to bacterial populations (29). In fact, the degree to which different bacterial species differ from panmixia is a subject of ongoing controversy and is one of the keys to a better understanding of bacterial population genetics. In principle, smaller numbers of gene copies and more clonal (further from panmixia) populations should lead to more inter-cistronic and intraspecific variation and, in the absence of strong purifying selection (such as those predicted for spacer regions), to faster evolutionary rates. Indeed, the levels of inter-cistronic heterogeneity and intraspecific variation seem to be higher in bacteria. While as far as we know, *E. coli* strains have the same number of copies of the *rm* operons, different situations are apparent in different strains. For example, *E. coli* K-12 and ECOR 40 are remarkably heterogeneous at the level of the entire operon. In contrast, strains in MLEE cluster B2, which includes strains that cause urinary tract infections, are much more homogeneous, reflecting either their inability to recombine with DNA from distantly related strains or a restricted ecological niche.

REFERENCES

1. Adoutte, A., and H. Philippe. 1993. The major lines of metazoan evolution: summary of traditional evidence and lessons from ribosomal RNA sequence analysis. *EXS* 63:1–30.
2. Antón, A. I., A. J. Martínez-Murcia, and F. Rodríguez-Valera. 1998. Sequence diversity in the 16S-23S intergenic spacer region (ISR) of the rRNA operons in representatives of the *Escherichia coli* ECOR collection. *J. Mol. Evol.* 47:62–72.
3. Brosius, J., T. J. Dull, D. D. Sleeter, and H. F. Noller. 1981. Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli*. *J. Mol. Biol.* 148:107–127.
4. Burgin, A. B., K. Parodos, D. J. Lane, and N. R. Pace. 1990. The excision of intervening sequences from *Salmonella* 23S ribosomal RNA. *Cell* 60:405–414.
5. Cilia, V., B. Lafay, and R. Christen. 1996. Sequence heterogeneities among 16S ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level. *Mol. Biol. Evol.* 13:451–461.
6. Condon, C., D. Liveris, C. Squires, I. Schwartz, and C. L. Squires. 1995. rRNA operon multiplicity in *Escherichia coli* and the physiological implications of *rml* inactivation. *J. Bacteriol.* 177:4152–4156.
7. Dover, G. A. 1992. Observing development through evolutionary eyes: a practical approach. *Bioessays* 14:281–287.
8. Dover, G. A., and R. B. Flavell. 1984. Molecular coevolution: DNA divergence and the maintenance of function. *Cell* 38:622–623.
9. DuBose, R. F., D. E. Dykhuizen, and D. L. Hartl. 1988. Genetic exchange among natural isolates of bacteria: recombination within the *phoA* gene of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 85:7036–7040.
10. García-Martínez, J., A. J. Martínez-Murcia, A. I. Antón, and F. Rodríguez-Valera. 1996. Comparison of the small 16S to 23S intergenic spacer region (ISR) of the rRNA operons of some *Escherichia coli* strains of the ECOR collection and *E. coli* K-12. *J. Bacteriol.* 178:6374–6377.
11. Gray, M. W., and M. N. Schnare. 1990. Evolution of the modular structure of rRNA, p. 589–597. In W. E. Hill, A. Dahlberg, R. A. Garrett, P. B. Moore, D. Schlessinger, and J. R. Warner (ed.), *The ribosome: structure, function, & evolution*. American Society for Microbiology, Washington, D.C.
12. Gurtler, V., and V. A. Stanisich. 1996. New approaches to typing and identification of bacteria using the 16S-23S rDNA spacer region. *Microbiology* 142:3–16.
13. Gutell, R. R. 1994. Collection of small subunit (16S- and 16S-like) ribosomal RNA structures: 1994. *Nucleic Acids Res.* 22:3502–3507.
14. Gutell, R. R., N. Larsen, and C. R. Woese. 1994. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol. Rev.* 58:10–26.
15. Hancock, J. M., D. Tautz, and G. A. Dover. 1988. Evolution of the secondary structures and compensatory mutations of the ribosomal RNAs of *Drosophila melanogaster*. *Mol. Biol. Evol.* 5:393–414.
16. Herzer, P. J., S. Inouye, M. Inouye, and T. S. Whittam. 1990. Phylogenetic distribution of branches RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. *J. Bacteriol.* 172:6175–6181.
17. Hillis, D. M., and S. K. Davis. 1988. Ribosomal DNA: intraspecific polymorphism, concerted evolution, and phylogeny reconstruction. *Syst. Zool.* 37:63–66.
18. Hillis, D. M., C. Moritz, C. A. Porter, and R. J. Baker. 1991. Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science* 251:308–310.
19. Koonin, E. V., and M. Y. Galperin. 1997. Prokaryotic genomes: the emerging paradigm of genome-based microbiology. *Curr. Opin. Genet. Dev.* 7:757–763.
20. Larsen, N. 1992. Higher order interactions in 23S rRNA. *Proc. Natl. Acad. Sci. USA* 89:5044–5048.
21. Ludwig, W., R. Rosello-Mora, R. Aznar, S. Klugbauer, S. Spring, K. Reetz, C. Beinfuhr, E. Brockmann, G. Kirchhof, S. Dorn, M. Bachleitner, N. Klugbauer, N. Springer, D. Lane, R. Nietupsky, M. Weizenegger, and K. H. Schleifer. 1995. Comparative sequence analysis of 23S rRNA from *Proteobacteria*. *Syst. Appl. Microbiol.* 18:164–188.
22. Luz, S. P., F. Rodríguez-Valera, R. Lan, and P. R. Reeves. 1998. Variation of the ribosomal operon 16S-23S gene spacer region in representatives of *Salmonella enterica* subspecies. *J. Bacteriol.* 180:2144–2151.
23. Martínez-Murcia, A. J., A. I. Antón, and F. Rodríguez-Valera. Patterns of sequence variation in two regions of the 16S rRNA multigene family of *Escherichia coli*. *Int. J. Syst. Bacteriol.*, in press.
24. Milkman, R. 1996. Recombinational exchange among clonal populations, p. 2663–2684. In F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli and Salmonella: cellular and molecular biology*, vol. 2. ASM Press, Washington, D.C.
25. Ochman, H., and R. K. Selander. 1984. Standard reference strains of *Escherichia coli* from natural populations. *J. Bacteriol.* 157:690–693.
26. Pennisi, E. 1998. Genome data shake tree of life. *Science* 280:672–674.
27. Pupo, G. M., D. K. R. Karaolis, R. Lan, and P. R. Reeves. 1997. Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains

- inferred from multilocus enzyme electrophoresis and *mdh* sequence studies. *Infect. Immun.* **65**:2685–2692.
28. **Selander, R. K., J. Li, E. F. Boyd, F.-S. Wang, and K. Nelson.** 1994. DNA sequence analysis of the genetic structure of populations of *Salmonella enterica* and *Escherichia coli*, p. 17–49. *In* F. G. Priest et al. (ed.), *Bacterial diversity and systematics*. Plenum Press, New York, N.Y.
 29. **Smith, J. M.** 1996. Population genetics: an introduction, p. 2685–2690. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, vol. 2. ASM Press, Washington, D.C.
 30. **Sogin, M.** 1991. Early evolution and the origin of eukaryotes. *Curr. Opin. Genet. Dev.* **1**:457–463.
 31. **Tautz, D., C. Tautz, D. Webb, and G. A. Dover.** 1987. Evolutionary divergence of promoters and spacers in the rDNA family of four *Drosophila* species. Implications for molecular coevolution in multigene families. *J. Mol. Biol.* **195**:525–542.
 32. **Woese, C. R.** 1987. Bacterial evolution. *Microbiol. Rev.* **51**:221–271.
 33. **Woese, C. R.** 1998. The universal ancestor. *Proc. Natl. Acad. Sci. USA* **95**:6854–6859.

AUTHOR'S CORRECTION

Intraspecific Diversity of the 23S rRNA Gene and the Spacer Region Downstream in *Escherichia coli*

ANA I. ANTÓN, ANTONIO J. MARTÍNEZ-MURCIA, AND FRANCISCO RODRÍGUEZ-VALERA

*Unidad de Microbiología, Centro de Biología Molecular y Celular, Universidad Miguel Hernández,
Campus de San Juan, 03550 San Juan, Alicante, Spain*

Volume 181, no. 9, p. 2703–2709, 1999. Page 2708: Acknowledgments should have been included as shown below.

Financial support was from grants CICYT PM95-0111 and DGICYT
PB96-0793-C04-02 of the Spanish Ministry of Education and Science.
We thank Stuart Ingham for illustrations.