

Variation of the Ribosomal Operon 16S-23S Gene Spacer Region in Representatives of *Salmonella enterica* Subspecies

SARA PÉREZ LUZ,¹ FRANCISCO RODRÍGUEZ-VALERA,^{1*} RUITING LAN,² AND PETER R. REEVES²

*Departamento de Genética y Microbiología, Campus de San Juan, Universidad de Alicante, 03080 Alicante, Spain,¹
and Department of Microbiology, The University of Sydney, Sydney, New South Wales, Australia²*

Received 13 November 1997/Accepted 16 February 1998

The 16S-23S spacer regions of two ribosomal operons (*rrnA* and *rrnE*) have been sequenced in seven representatives of the *Salmonella enterica* subspecies. Isolated nucleotide substitutions were found at the same sites as in *Escherichia coli* but the number of polymorphic sites was much larger, as could be expected for a more heterogeneous species. Still, as in *E. coli*, most of the variation found was due to insertions and/or deletions affecting blocks of nucleotides generally located at equivalent regions of the putative secondary structure for both species. Isolated polymorphic sites generated phylogenetic trees generally consistent with the subspecies structure and the accepted relationships among the subspecies. However, the sequences of *rrnE* put subspecies I closer to *E. coli* K-12 than to the other *S. enterica* subspecies. The distribution of polymorphisms affecting blocks of nucleotides was much more random, and the presence of equivalent sequences in distantly related subspecies, and even in *E. coli*, could reflect relatively frequent horizontal transfer. The smallest 16S-23S spacers in other genera of the family *Enterobacteriaceae* were also sequenced. As expected, the level of variation was much larger. Still, the phylogenetic tree inferred is consistent with those of 16S rRNA or housekeeping genes.

Ribosomal operons have acquired paramount relevance for the study of bacterial evolution and phylogeny. The rRNA 16S and 23S genes are the most widely used molecular chronometers for inferring microbial phylogeny and have been instrumental in developing a comprehensive view of microbial phylogeny and systematics (14). In most prokaryotes the ribosomal genes form an operon with the order 16S-23S-5S and are cotranscribed in a single polycistronic RNA that has to be processed to give the RNA species present in the mature ribosome (8). The spacer between the 16S and 23S genes is often called the internal spacer region (ISR). These spacers contain regions with secondary structures (6) and sometimes, very often in gram-negative bacteria, tRNA genes. The number of ribosomal operons in bacteria varies between 1 and 11 (13). When there is more than one it is common that they are not identical (intercistronic heterogeneity), particularly regarding the ISR sequence (7, 9, 30). In *Escherichia coli* K-12 and *Salmonella enterica* LT2 there are seven operons located at equivalent loci in the genetic maps of the chromosomes of each strain (21). In three of them the ISR contains two tRNA genes (ISR2) for isoleucine and alanine (operons A, D, and H in K-12 and A, D, and B in LT2) with an average size of about 450 bp. The remaining four (operons B, C, E, and G in K-12 and C, E, G, and H in LT2) have a single tRNA gene for glutamic acid (ISR1) and are consequently smaller (about 350 bp) (9).

The role of the spacer regions in the maturation of the polycistronic rRNA is far from clear. Some regions have secondary structures with stretches of canonical base pairing that are substrates for RNase III cuts and separation of the gene products (26, 29). However, the final stages in maturation, which include the trimming of the rRNAs to their final size, take place in the assembling ribosomal subunit with ribonucleoproteins as substrate. Here the interaction among the precursor rRNA (containing stretches of the spacer still at-

tached) and the ribosomal proteins seems to be essential for proper processing.

The variation found among relatively close taxa is known to be very high for the spacers of the rRNA operons (27). Even within the γ subclass of the class *Proteobacteria*, some species show hardly any similarity to *E. coli*. For example, *Haemophilus influenzae* has only about 22 nucleotides centered around the RNase III recognition site 3' to the 16S rRNA gene and about 12 nucleotides at the equivalent site 5' to the 23S gene which are highly similar to those of *E. coli*. The rest of the spacer (except for the tRNA gene and the antiterminator called box A [4]) is totally different in the two species. In species of *Pseudomonas* only the last two sequences remain similar (13). The extreme divergence in size and sequence of the spacers among different groups of prokaryotes, together with their location between highly conserved rRNA genes, makes them good markers for bacterial classification, and their use in rapid, and even automatic, identification has been suggested (3, 11, 13, 17, 20).

Recently (2), the ISR sequences of the seven operons of 12 representative strains of the *E. coli* reference (ECOR) collection were studied, a group of strains encompassing much of the phylogenetic diversity within this species (28, 31). The variation was, however, very restricted. In fact, at the level of single nucleotide substitutions, the degree of variation found for these spacer regions is not very different from that found for the 16S rRNA genes of the same strains (7, 23). The main source of heterogeneity (both intercistronic and interstrain) was the presence of insertions and/or deletions involving blocks of nucleotides. And most of these were already present in K-12 as intercistronic heterogeneity. There was a clear tendency toward homogenization (less intercistronic heterogeneity) in most strains. However, some apparently infrequent variations, such as the *rsI* sequence (6), maintained a widespread distribution, perhaps reflecting a horizontal transfer and recombination that counteracts the homogenization trend. These results represent an interesting problem of molecular evolution. First, if this sequence has indeed little functional restriction, as the great intergenic variation seems to indicate, greater

* Corresponding author. Present address: Unidad de Microbiología, Centro de Biología Molecular y Celular, Campus de San Juan, Apartado 18, 03550 San Juan, Alicante, Spain. Phone: 34 6 5919451. Fax: 34 6 5919457. E-mail: FRVALERA@UMH.ES.

variation should be expected. And second, there seems to be a sharp change in the rate of variation within and between species that contrasts with the relatively smooth gradient of variation found for the ribosomal genes.

To improve our understanding of the molecular evolution of these spacer regions, we have studied their sequence in representatives of *S. enterica* subspecies. Besides the location and tRNA genes present (21) nothing was known about the *S. enterica* ISRs. After *E. coli*, *S. enterica* is probably the bacterial species best known at the level of intraspecies variation and phylogeny (32). The structure of the subspecies seems to be very robust; very diverse types of data, from multilocus enzyme electrophoresis to sequencing of various genes, give results coherent with the accepted subspecies structure (5). Sequence data also point to a very preserved clonal structure within this species. We have also sequenced some tRNA^{Glu}-containing spacers of other members of the family *Enterobacteriaceae* to scan the variation found for wider phylogenetic gaps.

MATERIALS AND METHODS

Strains. The *S. enterica* strains utilized were obtained from different sources as described previously (35). The *S. enterica* subspecies and strains used were as follows: subspecies I, LT2 and M298; subspecies II, M495; subspecies IIIa, M313; subspecies IIb, M316; subspecies IV, M318 and M320; subspecies V, M321; and subspecies VI, M324. *Citrobacter freundii* OS60 (19), *Enterobacter aerogenes* E482 (19), *Klebsiella pneumoniae* LD119 (19), *Escherichia blattae* ATCC 29907, *Escherichia hermannii* ATCC 33650, *Escherichia vulneris* ATCC 33821, and *Escherichia fergusonii* ATCC 35469 were kindly provided by J. G. Lawrence (Washington University).

PCR primers. To specifically amplify the ISRs of operons *rmA* and *rmE*, two primers were designed located in genes upstream of both operons in *E. coli* K-12. For *rmA* the primer TCTCAACAAGGACGGACAA, which corresponds to nucleotide positions 20 to 39 in *hemG* of *E. coli* K-12, was used. For *rmE* the primer used was TAATGGCATCGTCTGGCCGC. This primer is derived from positions 1322 to 1342 of *purH*. Universal primers 16S14F and 23S1R have been described elsewhere (12).

DNA extraction and PCR amplification. One colony from a fresh culture was resuspended in 200 μ l of Instagene Purification Matrix (Bio-Rad) in Eppendorf tubes. DNA was extracted following the manufacturer's recommendations, and the DNA concentration was estimated with a Gene Quant RNA/DNA calculator (Pharmacia LKB Biochrom). For PCR the DNA concentration was adjusted to 20 ng/ μ l. Amplification of specific *rm* operons in *S. enterica* strains was carried out in two steps (nested PCR). Approximately 40 ng of chromosomal DNA was subjected to PCR amplification in a total volume of 50 μ l containing 50 mM KCl, 10 mM Tris-HCl (pH 9.0), 1.5 mM MgCl₂, 0.1% Triton X-100, 0.2 mM (each) deoxyribonucleotide (dATP, dCTP, dGTP, and dTTP; Pharmacia LKB Biotechnology), 2 U of *Taq*Plus DNA polymerase (Stratagene), and 50 to 60 pmol of each primer. For the first step, *rm*-specific primers were used in combination with 23S1R. The reaction mixtures were subjected to a "touch down" (10) thermal cycling regime which consisted of the following: 3 cycles of 94°C for 15 s, 65°C for 30 s, and 72°C for 3 min; 3 cycles of 94°C for 15 s, 62°C for 30 s, and 72°C for 3 min; 3 cycles of 94°C for 15 s, 59°C for 30 s, and 72°C for 3 min; 3 cycles of 94°C for 15 s, 56°C for 30 s, and 72°C for 3 min; and 23 cycles of 94°C for 15 s, 55°C for 30 s, and 72°C for 3 min; followed by an extension step at 72°C for 5 min. The amplicon from this reaction was diluted 1:10,000 and submitted again to PCR with the universal primers 16S14F and 23S1R as follows: 35 cycles of 94°C for 15 s, 62°C for 30 s, and 72°C for 1 min plus an extension step of 5 min at 72°C. For all other strains of *Enterobacteriaceae* only the second PCR protocol was carried out followed by cloning. All reactions were carried out in a PTC-100 Peltier effect thermocycler (MJ Research Inc.). The PCR products were separated in 1% agarose gels in Tris-acetate-EDTA buffer, stained with 0.5 μ g of ethidium bromide, and visualized with UV.

Purification of the PCR products. PCR products were purified with the QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's protocol. The DNA was recovered in 30 μ l of water and diluted to a concentration between 0.3 and 0.5 μ g/ μ l.

Cloning of *Enterobacteriaceae* PCR products. For the *Enterobacteriaceae* other than *S. enterica* no operon-specific primers were used, and therefore all the operons were amplified simultaneously. The PCR product was cloned using the original TA Cloning Kit (Invitrogen) following the manufacturer's recommendations. From each PCR product 20 colonies were picked and grown in Luria Bertani medium at 37°C for 18 h. Recombinant plasmids were extracted by using the QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's recommendations. The purified plasmids were digested with *Eco*RI to separate the insert, and the product was run in agarose gels to determine insert size. For each strain one clone containing the smallest insert was selected for sequencing.

Sequencing of 16S-23S ISR. Nucleotide sequences of PCR products and plasmids were determined by using the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer) and an ABI PRISM 377 sequencer (Perkin-Elmer) according to the manufacturer's indications. The 16S14F and 23S1R primers were used as sequencing primers in all cases. For every PCR product both strains were sequenced and every sequence was repeated at least twice.

Data analysis. Sequences were aligned by using the PCGene program (IntelliGenetics Inc.). All other analyses were done using the MEGA (Molecular Evolutionary Genetics Analysis) 1.01 program obtained from the Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, University Park, Pa.

Nucleotide sequence accession numbers. *S. enterica* sequences determined here were assigned GenBank accession numbers AFO45563 and AFO46806 to AFO46822; representatives of other *Enterobacteriaceae* were assigned accession numbers AFO47420 to AFO47426.

RESULTS

We have sequenced the *S. enterica* *rmA* and *rmE* operons as defined by the upstream sequences found in *E. coli* K-12. However, except for LT2, we have no direct evidence about the location of the operons in the genomic map of the strains studied. In fact, for *Salmonella typhi* a highly plastic chromosomal organization has been previously described (22). In *E. coli* strains the general organization of the *rm* operons in the genome seems to be quite conserved. Not only is the combination of four ISR1- and three ISR2-containing operons kept throughout the ECOR collection but the type of operon (as defined by the tRNA genes in the ISR) present at each locus is also highly conserved (2). However, in *S. enterica* there seems to be variation at this level. LT2 has an ISR1 present in *rmE* and an ISR2 in *rmA*. However, in M313 (subspecies IIIa) and M318 and M320 (both subspecies IV) the primer for *rmA* gave operons with ISR1 present. It is possible that, in these three strains, a two-tRNA-containing operon was present at another locus to keep the four ISR1- and three ISR2-containing operon distribution found in *E. coli* and LT2. However, in all these strains the operon present at the *rmE* locus was an ISR1 as well. When an ISR1 was retrieved with the primer for *rmA* the sequence found was very similar to the *rmE* sequence determined for the same strain. Specifically, in M313 and M318 the two ISR1 sequences differed by only 6 and 1 nucleotides, respectively. In M320 the ISR1 obtained with the primer for *rmA* differed in 4 aligned nucleotides but lacked two of the insertions found in the ISR1 of *rmE* (see below). That does not necessarily mean that the ISR1 spacers present at the putative *rmA* locus derive from the *rmE* operon, since, at least in *E. coli*, intercistronic heterogeneity in many strains is low (2) and any of the ISR1-containing operons could have a sequence very similar to that of *rmE*. To simplify the terminology the ISRs obtained with the *rmE* and *rmA* primers are identified in the text and figures by the letters E and A, respectively, and the strain designation.

Variability among the ISR1 operons. The alignment of the ISR1 operons (containing a single tRNA gene) is shown in Fig. 1. Figure 1 includes the nine sequences retrieved from the strains with the primers designed for *rmE* as well as three sequences obtained with the primers for operon *rmA*. Two types of sequence variation were found. One variation affected aligned residues only and was used to analyze level of divergence and to infer relationships. The other involves insertions and/or deletions or block substitutions and will be discussed separately (see below). The polymorphic nucleotides of ISR1 and their relation to the secondary structure are shown in Fig. 2. In the *S. enterica* ISR1s that were retrieved the maximum pairwise nucleotide divergence was 18% and the average divergence was 9.8%. This is about three times higher than that found for the *rmE* operon in *E. coli* (2.6%) (2), perhaps re-

ISR1

	.10	.20	.30	.40	.50	.60	.70	.80
<i>rrnEK12</i>	CCTTAAGAA	GGTTCCTTG	CAGTGCTCAC	ACAGATTGTC	TGATAGAAAG	TGAA-AAGCA	AGGCGTCTTG	CGAAGCAGAC
ELT2 (I)		...A....	A.....		...GA....	...T....		...TTG...
E298 (I)	...G....	A...A....			...GA....	A...T....		...A....
E313 (IIIa)	...G....	C...G....			...GA....	A...C...T..		...TAG.A..
A313 (IIIa)	...G....	A...G....			...GA....	AG.G....		...GAG....
E316 (IIIb)	...G....	A...G....			...GA....			...TTG....
E318 (IV)	..AG....	C...G....			...GA....			...TTG....
A318 (IV)	..AG....	C...G....			...GA....			...TTG.T..
E320 (IV)	...G....	C.TGC....	T...TC....		...TGT....	A...G....		...TTG....
A320 (IV)	...G....	C.TGC....	T...TC....		...A.TGT....	A...G....		...TTG....
E321 (V)	...G....	CT.....			...GA....	A...T....		...TTG....
E324 (VI)	...G....	...A....	A.....		...GA....	A...T....		...TTG....
E495 (II)	...G....	...A....			...GA....	A...T....		...TTG....
	** *****		*****	*****	*****			
			.170	.180	.190	.200	.210	.220
<i>rrnEK12</i>	TGATAC (tRNA Glu)	-----	-----	-----	-----	-CTTGCTGGT	TTGTGAGTGA	
ELT2 (I)	.TCAGT	"	GGGTTCAA	ACTGATGAGG	TCAAACCTCC	AGGGGACGCC	A.....	
E298 (I)		"						
E313 (IIIa)	.AT-GA	"					GC.G	AA..T..
A313 (IIIa)	.AT-GA	"					GC.G	AA..T..
E316 (IIIb)	.TCAGT	"						
E318 (IV)	CAT-GA	"	GGGTTCAA	ACTGATGAGG	TCAAACCTCC	AGGGGACGCC	A.....	
A318 (IV)	CAT-GA	"	GGGTTCAA	ACTGATGAGG	TCAAACCTCC	AGGGGACGCC	A.....	
E320 (IV)	.TCAGT	"	GGGTTCAA	ACTGATGAGG	TCAAACCTCC	AGGGGACGCC	A.....	
A320 (IV)	.TCAGT	"				A.....		
E321 (V)	.TCAGT	"					GC.G	AA..T..
E324 (VI)	.TCAGT	"						
E495 (II)	.TCAGT	"						
			.230	.240	.250	.260	.270	.280
<i>rrnEK12</i>	AAGTCACCTG	CCTT-AATAT	CTCAAACTC	AT-----	-----CTTCG	GGTGATGTTT	SAGATATTG	CICITTAATA
ELT2 (I)			G		T.A..	A..C.C....		
E298 (I)			G		T.A..	A..C.C....		
E313 (IIIa)	...G.GT...	...A.C.G...	G		T.A..	A..C.C....		
A313 (IIIa)	...G.GT...	...A.C.G...	G		T.A..	A..C.C....		
E316 (IIIb)	G	G	G		T.A..	A..C.C....		
E318 (IV)	G	G	G		T.A..	A..C.C....		
A318 (IV)	G	G	G		T.A..	A..C.C....		
E320 (IV)	G	G	G		T.A..	A..C.C....		
A320 (IV)	G	G	G		T.A..	A..C.C....		
E321 (V)	...G.GT...	...A.C.G...	G		T.A..	A..C.C....		
E324 (VI)	...G.GT...	...A.C.G...	G		T.A..	A..C.C....		
E495 (II)	G	G	G		T.A..	A..C.C....		
			.310	.320	.330	.340	.350	.360
<i>rrnEK12</i>	ATCTGGATCA	AGCTGAAAT	TGAACACTG	AACAACGAA	G--TTCTTCG	TGAGTCTCTC	AAATTTTCGC	AACACGATGA
ELT2 (I)			A		G....			
E298 (I)			A		G....			
E313 (IIIa)			A	AT..	G....	G....	T..	...GT..A..
A313 (IIIa)			A	AT..	G....	G....	T..	...GT..A..
E316 (IIIb)			A	TTCCG	A...G...A			...TCATGAAG
E318 (IV)			A	T.T..	CGA..G....			...TCA..A..
A318 (IV)			A	T.T..	CGA..G....			...TCA..A..
E320 (IV)			A	T.T..	CGA..G....			...TCA..A..
A320 (IV)			A	T.T..	CGA..G....			...TCA..A..
E321 (V)			GACA	C...GA...T	GTG..	C.GA....		...TCA..A..
E324 (VI)			A	T.TT	A...G...A			...TCATGAAG
E495 (II)			A	TTTCC	GA...G...A			...TCA-GAAG
			.390	.400	.410	.420	.430	.440
<i>rrnEK12</i>	TGAATCGTAA	GAA-----	-----	-----	-----	-----	-----	-----
ELT2 (I)								
E298 (I)	AC..							
E313 (IIIa)								
A313 (IIIa)								
E316 (IIIb)	ATG...TGT	A...GGTG...	AAGTGAGCGA	AGTGAAGACA	AGGCGTACCG	CACGCAGCGA	GCGCAGTGGA	CTTATGTC
E318 (IV)								
A318 (IV)								
E320 (IV)								
A320 (IV)								
E321 (V)								
E324 (VI)								
E495 (II)	AC...TGT	A...GGTG...	AAGTGAGCGA	AGTGAAGACA	AGGCGTACCG	CACGCAGCGA	GCGCAGTGGA	CTTATGTC

FIG. 1. Alignment of the nucleotide sequences of ISR1 and ISR2 spacers. Letters (A and E) indicate the specific primers (for *rmA* and *rmE*, respectively) used for the PCR amplification of the sequences. The numbers indicate the strains, and the roman numerals in parentheses indicate the subspecies to which each strain belongs. K-12 *rmE* was used as the reference strain. Nucleotides conserved in all *Enterobacteriaceae* studied here are labelled with an asterisk. The variable region between tRNA^{Ile} and tRNA^{Ala} of the ISR2 spacers is shown as shaded or in boldface or italics to indicate each of the three types of sequence found. Box A is underlined.

flecting the wider phylogenetic span (genetic variability) included in *S. enterica* (25). Although the number of polymorphic sites is much larger in the ISR1 spacers of *S. enterica*, the variation in this region, similar to that of *E. coli*, is rather restricted and is concentrated at some hypervariable sites, including those already detected in *E. coli*. For the 70 nucleotides downstream of the 16S rRNA gene, in addition to the 8 polymorphic sites which are also present in *E. coli* (2), *S. enterica* strains have 14 additional polymorphic sites (a total of 22 polymorphic sites).

A phylogenetic tree constructed by the neighbor-joining method and comparing the ISR1 sequences retrieved is shown in Fig. 3. The main clusters reflect the subspecies structure and are generally concordant with a tree of three housekeeping

(HK) genes (5). Subspecies I, VI, II, IV, and IIIb form a cluster as do the five HK genes studied by Boyd et al. (5), and subspecies V and IIIa appear clearly distant from the rest, which is also the case for the HK gene tree. Some sequences found in *S. enterica*, such as those of E298 and ELT2, are more similar to *rmE* of *E. coli* K-12 than they are to other *S. enterica* strains.

As found in *E. coli* *rm* operons, there are polymorphisms clustered in hypervariable regions in which substitutions affecting blocks of nucleotides occur, sometimes with insertions and deletions of highly variable lengths. At positions 66 to 86 (Fig. 1 and 2), some *E. coli* ISR1-containing operons present the *rsI* sequence (6), an insertion of 106 nucleotides. In the *S. enterica* strains studied the last 10 nucleotides are highly polymorphic, although all the alternative sequences found were of equivalent

		.470	.480	.490	.500	.510	.520	.530	.540
EK12	-----	-----	-----	-----	-----	-----	---ACATCTT	CGGGTTGTGA	
ELT2 (I)									
E298 (I)									
E313 (IIIa)									
A313 (IIIa)									
E316 (IIb)	ATGAGCATCG	CGAGTACGGA	ACAACGCAGT	ATTACACACG	CGCAACAAC	CGGACAGAAC	AAG.....		
E318 (IV)									
A318 (IV)									
E320 (IV)									
A320 (IV)									
E321 (V)									
E324 (VI)									
E495 (II)	ATGAGCATCG	CGAGTACGGA	ACAACGCAGT	ATTACACGATG	CGCAACAACG	CGGACAGAAC	AAG.....	*****	

		.10	.20	.30	.40	.50	.60	.70	
K12	CCTTAAGAA	GCCTACTTTG	CAGTCTCAC	ACAGATTGTC	TGATGAAAA	-GAGCAGTAA	AACCTCTACA	(tRNA Ile)	
LT2 (I)		A..GT.....	A.....		A	C.....		"	
298 (I)		A..GT.....			A	C.....		"	
316 (IIb)		A..GT.....			A	C.....		"	
321 (V)		CT..T.....			A	C.....		"	
324 (VI)		A..GT.....			A	C.....		"	
495 (II)		A..GT.....			A	C.....		"	

		.150	.160	.170	.180	.190	.200	.210	.220
K12	AAATT	TGCAC----	GGCAAAATTG	AAGAG-----	-----	GTTTAACTA	CATGTA--	-----	
LT2 (I)		T..CTGAA	TA.TGCG...	TGA.ATAACT	CACATACTG	A.G.ATG..T	G.TA.TCCA	CGCCTTGCT	
298 (I)		T..CTGAA	TA.TGCG...	TGA.ATAACT	CACATACTG	A.G.ATG..T	G.TA.TCCA	CGCCTTGCT	
316 (IIb)		C..T.CCG..	T..TTTG...	TGGCAAGCT	CGCATACTT	AAG.ATG..T	GCT...CCA	CGCCGCGCC	
321 (V)		T..CTGAA	TA.TGCG...	TGA.ATAACT	CACATACTG	A.G.ATG..T	G.TA.TCCA	CGCCTTGCT	
324 (VI)		TT-ACTGA	T..TGCG...	CGCGACACT	CACATACTT	.AG.ATG..T	G.TA.TCCA	CGCCTTGCT	
495 (II)		T..CTGAA	TA.TGCG...	TGA.ATAACT	CACATACTG	A.G.ATG..T	G.TA.TCCA	CGCCTTGCT	

		.230	.240	.250	.260	.340	.350	.360	.370
K12	-----	-----	-----	-----	-----	T (tRNA Ala)	T -CTCTGTAGT	GGTTAAATA	AAAACTCTC
LT2 (I)	CAGGAAAAAT	TATCGGTAAA	GAGGTTTC..T	GACTACACGA	"	GTG...	T..T.CG..		
298 (I)	CAGGAAAAAT	TATCGGTAAA	GAGGTTTC..T	GACTACACGA	"	GTG...	T..T.CG..		
316 (IIb)	CGGAAACGAA	TCAAAGGTGA	GAGGTTCCCT	GACTACACGA	A.	GTG...	T..T.CG..		
321 (V)	CAGGAAAAAT	TATCGGTAAA	GAGGTTTC..T	GACTACACGA	"	GTG...	TA..T.CG..		
324 (VI)	CAGGAAAAAT	TATCGGTAAA	GAGGTTTC..T	GACTACACGA	"	A..G.AG.AC	TAAT-CA.G		
495 (II)	CAGGAAAAAT	TATCGGTAAA	GAGGTTTC..T	GACTACACGA	"	GTG...	T..T.CG..		

		.380	.390	.400	Box A .410	.420	.430	.440	.450
K12	AGAGTGTACT	TGCAAGGTTT	-CACTGCGAA	GTITIGCTCT	TTAAAAATCT	GGATCAAGCT	GAAAAATGAA	ACACTGAACA	
LT2 (I)		A..G....	T.....		C.....			A..T	
298 (I)		A..G....						A..	
316 (IIb)		A..G....						A..	
321 (V)		A..G....						A..	
324 (VI)		A..G....						A..	
495 (II)		A..G....						A..	

		.460	.470	.480	.490	.500	.510	.520	.530
K12	ATGAAAGTGT	TT-CGTGAGT	CTCTCAAATT	TTCCGAACCT	TGAAGTGAA	-----	-----	-----	
LT2 (I)	C.....		A.....	A	GATGA...T	CGTAAGAA..	-----	-----	
298 (I)	C.....		A.....	A	GATGA...T	CGTAAGAA..	-----	-----	
316 (IIb)	C.....C...			G	GATGA...T	CGTAAGAA..	-----	-----	
321 (V)	..TCCGA...			TCA	ATG.C	CTGTAAAGG	TGAAGTGAGC	GAAGAGAAGA	
324 (VI)	..TT.A...	..T.A...		TCA		CTGTAAAGG	TGAAGTGAGC	GAAGAGAAGA	
495 (II)	..TCCGA...			TCA	GA.GA----				

		.540	.550	.560	.570	.580	.590	.600	.610
K12	-----	-----	-----	-----	-----	-----	-----	-----	
LT2 (I)									
298 (I)									
316 (IIb)									
321 (V)	CAAGCGGTAC	CGCACGCAGC	GAGCGCAGTG	GACTCGATGT	CCATGAGCAT	CGCGAGTACG	GAACAACGCA	GTATTCACAA	
324 (VI)									
495 (II)									

		.620	.630	.640	.650	
K12	-----	-----	---ACATC	TTGGGTTGT	GA	
LT2 (I)					..	
298 (I)					..	
316 (IIb)					..	
321 (V)	CGCGCAACRA	ACCGGACAGA	ACAAG.....		..	
324 (VI)					..	
495 (II)					..	

FIG. 1—Continued.

lent length, similar to that of the *E. coli* ISR1 without the *rsf* sequence. The second of such variable regions is located downstream from the tRNA^{Glu} gene and corresponds to an insertion found previously in *rmE* and *rmC* of ECOR 40 (2) which contains a perfect direct repeat of the last 11 bases of the tRNA^{Glu} gene (Fig. 2). Sequences A318, E318, E320 (all representatives of subspecies IV), and ELT2 (subspecies I) have this insertion. In nucleotides 234 to 274 there is a secondary structure loop (Fig. 2) that has been shown to be hypervariable in *E. coli* (2, 12, 16), in which species there are at least three versions which have little similarity to one another except that each maintains the same size and general structure of the loop (2). In the strains of *S. enterica* studied the version found in ECOR 35 and ECOR 40 predominates; E321 has a version that is almost identical to version I (2) of *E. coli*, which is

present in most operons and strains in this species. ISRs A318, E318, and E320 have yet another different and longer loop (box B in Fig. 2) that has not been found in *E. coli*. The secondary structure loop from nucleotide 307 to nucleotide 338 (Fig. 2) concentrates 18 polymorphisms with most strains diverging widely from the sequence found in *E. coli* and the two representatives of subspecies I. Finally, the region downstream from nucleotide 358 is also highly polymorphic. In *E. coli* there are two alternative versions of 17 and 8 nucleotides depending on the strains and operons (2). In *S. enterica* two ISRs, ELT2 and E298, had sequences very similar to the 17-mer version of *E. coli* (the most abundant in this species). Two ISRs, E495 and E316, had a large insertion of approximately 145 bp occupying this region. This insertion could form a long secondary structure loop (Fig. 2). The rest (E324, A313,

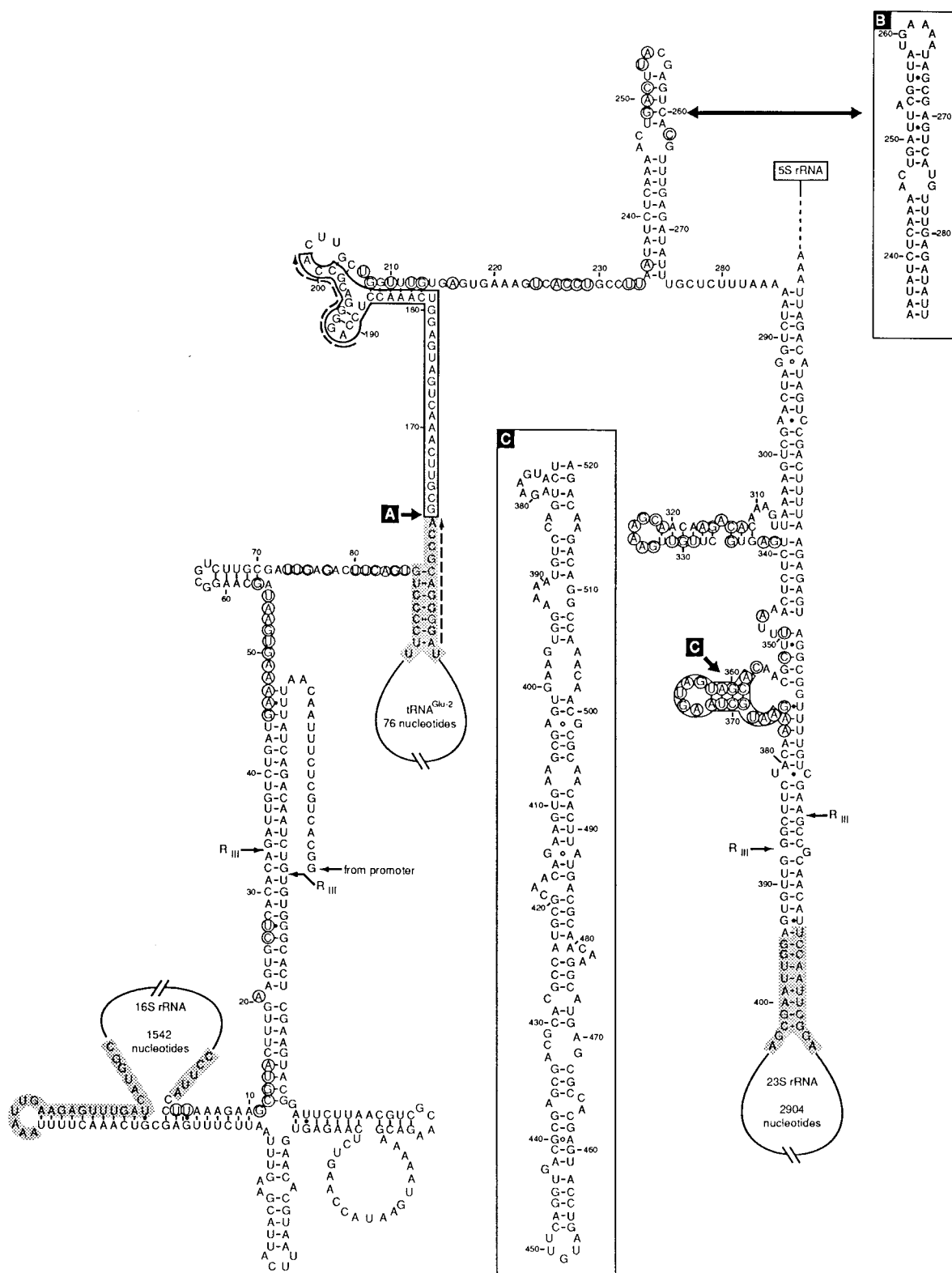


FIG. 2. Secondary structure of the *mrE* ISR of *S. enterica* LT2 based on the model suggested by Brosius et al. (6). Variable positions in the *S. enterica* strains studied are indicated by circles. Block substitutions and insertions are boxed. B and C boxes show alternative sequences found in some operons and strains and putative secondary structures. The ends of the 16S and 23S rRNAs and tRNA genes are shaded. RNase III cutting sites are indicated (R_{III}). Dashed arrows indicate pairs of direct repeats.

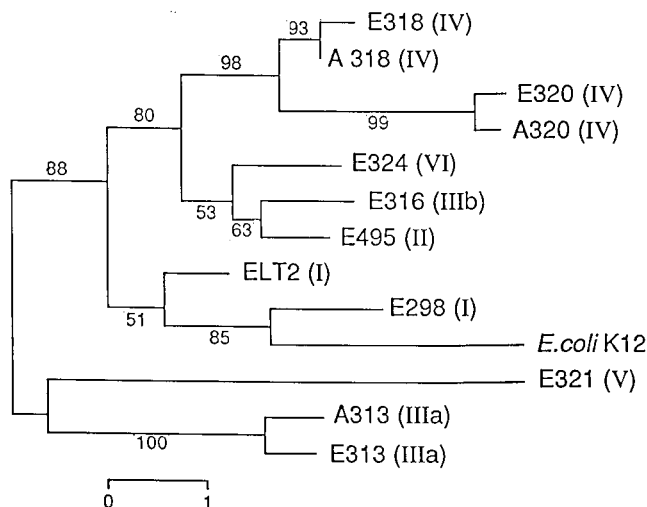


FIG. 3. Dendrogram based on ISR1 sequences showing the relationships among *E. coli* K-12 and representatives of *S. enterica* subspecies. The tree was constructed by the neighbor-joining method. Boot-strap values were calculated from 500 trees. Each number on a branch indicates the percentage of trees in which the node was supported. Bar, percent sequence divergence.

E313, E321, A318, E318, E320, and A320) all have a stretch of 10 bp vaguely reminiscent of the octamer found in a few strains of *E. coli*.

Most polymorphisms, particularly those affecting clusters of nucleotides, are located outside the regions that are involved in secondary structure with nucleotide stretches upstream or downstream of the 16S or 23S gene, respectively, in the transcript (Fig. 2). This is understandable considering the role of these regions as RNase III targets and the requirements for concerted variation with regions distantly located in the operon.

Variability among the ISR2. The *rrnA*-specific PCR primer generated amplicons with an ISR2 (containing tRNA^{Ile} and tRNA^{Ala} genes) only in six of the nine strains studied here. Therefore, the alignment of the sequences for the ISR2-containing operons shown in Fig. 1 comprises only 6 *S. enterica* sequences plus the sequence for the *E. coli* K-12 *rrnA* operon included for comparison. The ISR2-containing operons obtained have very little sequence diversity for the 150-odd bases that comprise the tRNA^{Ile} plus the 70 bp located upstream from this gene. This region is also very similar to the *E. coli* ISR2-containing operons. Downstream from the tRNA^{Ile}, variability is very high, and in fact, this stretch in all the *S. enterica* strains diverged considerably from that in *E. coli*, in which all the strains and operons studied have an identical sequence of 43 bp (2). In *S. enterica* this stretch, i.e., downstream from the tRNA^{Ile}, is much larger in all strains (about 120 bp). Among the six strains of *Salmonella* studied, four (LT2 and M298, both subspecies I; M495, subspecies II; and M321, subspecies V) had nearly identical sequences (one nucleotide different) of 112 bp. M324 (subspecies VI) and M316 (subspecies IIIb) had stretches of 111 and 114 bp, respectively, but with widely divergent sequences. The three types of sequence have little in common except overall length and the first 5 and last 11 nucleotides. A316 is different throughout the whole sequence, while A324 differs mostly in the first 50 nucleotides of this stretch, being for the last 60-odd bases identical to that of ALT2 and the other sequences. Downstream of the tRNA^{Ala} there is little heterogeneity at the level of isolated nucleotide substitutions, and again the sequence is rather similar to that of

E. coli. The hypervariable loop located near the 5' end of the 23S gene (Fig. 2, box C) is also present in the ISR2 spacers. Here the variability is extremely high to the point that none of the sequences determined was identical to any of the others. ALT2, A298, and A316 have 20-mer stretches that are highly similar to the 19-mer stretch found in *rrnE* in *E. coli* K-12. A321 has a long loop similar to that which was also found in this region in E316 and E495 (both ISR1). A495 has a very short stretch of 7 bp, again with a completely different sequence from that found in the *rrnA* operon of *E. coli* K-12. Finally, A324 has still another sequence of 11 bp, which is similar to the sequence present in *rrnA* of K-12.

ISR1 spacers in other Enterobacteriaceae. To get an impression of how the spacer sequence varies with increasing phylogenetic distance, we sequenced the smallest ISRs present in the ribosomal operons of seven strains of different species of *Enterobacteriaceae*. In fact operons belonging to *C. freundii* and *Enterobacter aerogenes* had spacers of 235 and 198 bp, respectively, that did not contain a tRNA gene and thus cannot be considered functionally equivalent to the other ribosomal operons. A tree which compares these sequences with those of *rrnE* of *E. coli* K-12 and *S. enterica* LT2 is shown in Fig. 4. The conserved positions for all the *Enterobacteriaceae* studied are shown in Fig. 1. Only the regions around the RNase III recognition site and box A are conserved for all of the species studied. Generally speaking, the clustering was similar to that developed from other gene comparisons (1, 19, 32). *Escherichia fergusonii* is a sister species to *E. coli*, and *S. enterica* appears as its closest relative in the ISR1 tree. These relationships are also evident in a comparison of HK genes. *K. pneumoniae* and *E. vulneris* form a cluster that also appears in a phylogenetic tree generated by comparing *gapA* and *ompA* genes (19). At the level of the ISR1 this cluster is already widely divergent. *Klebsiella* is the closest genus to *E. coli* and *Salmonella* by 16S rRNA sequence (1). *E. blattae* and *E. hermannii* are very divergent, which is consistent with the genus *Escherichia* being polyphyletic as has been shown by other authors (19).

DISCUSSION

The relative lack of functional restriction of the rRNA operon spacers would make us expect a rather high level of sequence variation. Still, some data in the literature hints at

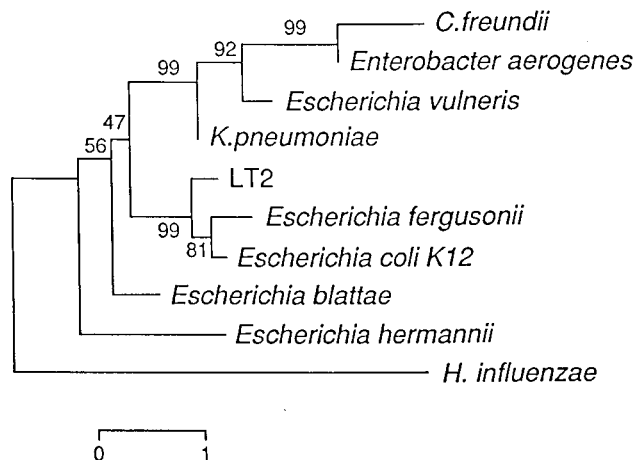


FIG. 4. Dendrogram based on the sequence of the smallest ISRs-present in the *rrn* operons of different *Enterobacteriaceae*. Computation and bar are as described in the legend to Fig. 3.

unexpected levels of conservation in other spacer sequences. For example, in the classical studies by Milkman and coworkers of the *trp* operon of *E. coli* (24, 34) there is information showing that the variation in the noncoding areas is similar to, or only slightly higher than, that in the coding ones; the main difference was that in the coding areas the variation seemed to concentrate on the third base of each codon (silent mutations). If the noncoding or spacer areas of the bacterial genomes really are devoid of function, something that is by no means certain, it could be considered an argument in favor of the neutral hypothesis for molecular evolution (18), i.e., if variation is found with similar frequencies in functional and nonfunctional regions it could be because most of it is neutral.

Altogether, the results of this work strengthen the view based on a previous work with *E. coli* (2) in the sense that, at the level of isolated nucleotide substitutions, the ISRs do not differ so much from the rRNA genes themselves. The main difference is the presence of relatively large insertions and/or deletions. These drastically alter the primary structure of the spacer sequence and in the long run can probably also affect the secondary structure and even the function of the different regions of the spacer DNA. This phenomenon is by no means exclusive to the ISR sequences. For example, the spacer between the genes *rpsU* and *dnaG* only varies in 15 of 108 alignable nucleotides between *E. coli* K-12 and *S. enterica* LT2 (33). However, in *S. enterica* there is an insertion of 120 bp that is not present in *E. coli*. Other bacterial species have been shown to differ due to insertion and/or deletion events affecting the central region of the *rm* spacer (12, 36). Since it appears that the size of the spacer must vary within certain limits, it is foreseeable that the accumulation of insertions could eventually lead to deletions affecting other areas and thus to the gradual substitution of the overall spacer sequence, as is found when relatively wide phylogenetic distances are analyzed at this level.

The phylogeny generated by the ISR1 sequences is roughly consistent with that of 16S rRNA or housekeeping genes in *Enterobacteriaceae*, as has been shown for other bacterial groups (20). However, the distribution of clustered polymorphisms did not form a coherent picture with any of the relationship trees inferred from HK sequences or multilocus enzyme electrophoresis. Even the barrier between the two main clusters of *S. enterica* subspecies, i.e., subspecies IIIa and V and subspecies I, II, IIIb, IV, and VI, appears inconsistent with the distribution of some of the polymorphisms. For example, the long loop of box C (Fig. 2) is found in M316 (subspecies IIIb) and M495 (subspecies II) and a similar loop is present in the ISR2 of M321 (subspecies V). The same applies to the sequence located between the two tRNA genes in ISR2 of A321 (subspecies V) that is also present in a representative of subspecies II (A495) and in those of subspecies I (LT2 and A298). The presence of some of these sequences in relatively separated lineages of *S. enterica* and even *E. coli* hints at a relatively widespread horizontal transfer and recombination within and between species. In a previous work (2) a model was suggested in which two opposing forces act upon the *rm* operons. Homogenization processes would lead to little intergenic heterogeneity and conservation of the operon sequence. On the other hand, horizontal transfer from other strains (that homogenize different versions of the operon sequence), followed by recombination, acts as a continuous source of variation and intergenic heterogeneity. The diversity found in the different operons and strains results from the equilibrium between these two opposed processes (2). In *S. enterica* recombination seems to be even more widespread, acting among different subspecies and perhaps even species (*E. coli*).

It is remarkable that in terms of the ISR1 sequences there is no sharp discrimination between *E. coli* and *S. enterica*, i.e., some sequences found in *S. enterica* are more similar to *E. coli* than they are to other *Salmonella* strains. The main difference between the two species was found in the inter-tRNA gene stretch of the ISR2 spacers. Otherwise it would be difficult to delimit both species as separate entities by ISR sequences. For example, E298 (subspecies I) has only 19 nucleotide differences from the *rmE* gene of K-12, which is much less than the number of nucleotide differences found for the *rmE* gene of some strains of *E. coli* (2) or those between the *rmE* of E289 and other ISR1s of *S. enterica* strains. As shown in Fig. 1 the subspecies II (M495) and subspecies I (M298 and LT2) representatives have, over most of the ISR1, an *E. coli*-like sequence, although LT2 and E495 have *Salmonella* features like the sequence downstream of TRNA^{Glu} found in ELT2 or the insertion of box C (Fig. 2) in E495. This mosaic-like structure is normally attributed to recombination, although the possibility of these features corresponding to ancestral traits cannot be totally disregarded.

On the other hand, the average (most frequent bases at polymorphic sites) sequence found in *S. enterica* is significantly different from the average sequence in *E. coli*. This can be interpreted as indicating that both species are linked by a continuum of diversity rather than having the sharp discontinuity expected for sexual and isolated populations (15).

ACKNOWLEDGMENTS

Financial support was from CICYT PM95-0111 to F.R.-V. and from the Australian Research Council to P.R.R. Part of the work was carried out while F.R.-V. was a Visiting Academic at the Department of Microbiology, The University of Sydney, supported by a Generalitat Valenciana scholarship.

We are grateful to K. Hernández for secretarial assistance and S. Ingham for graphics.

REFERENCES

- Ahmad, S. W. G. Weisburg, and R. A. Jensen. 1990. Evolution of aromatic amino acid biosynthesis and application to the fine-tuned phylogenetic positioning of enteric bacteria. *J. Bacteriol.* **172**:1051–1061.
- Antón, A. I., A. J. Martínez-Murcia, and F. Rodríguez-Valera. 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.*, in press.
- Barry, T., G. Collieran, M. Glennon, L. K. Dunican, and P. Gannon. 1991. The 16S/23S ribosomal spacer region as target for DNA probes to identify eubacteria. *PCR Methods Appl.* **1**:51–56.
- Berg, K. L., C. Squires, and C. L. Squires. 1989. Ribosomal RNA operon anti-termination function of leader and spacer region box B-box A sequences and their conservation in diverse microorganisms. *J. Mol. Biol.* **209**:345–358.
- Boyd, E. F., F.-S. Wang, T. S. Whittam, and R. K. Selander. 1996. Molecular genetic relationships of the salmonellae. *Appl. Environ. Microbiol.* **62**:804–808.
- 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.
- 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.
- Condon, C., J. Phillips, A. Fu, C. Squires, and C. L. Squires. 1992. Comparison of the expression of the seven ribosomal RNA operons in *Escherichia coli*. *EMBO J.* **11**:4175–4185.
- Condon, C., C. Squires, and C. L. Squires. 1995. Control of rRNA transcription in *Escherichia coli*. *Microbiol. Rev.* **59**:623–645.
- Don, R. H., P. T. Cox, B. J. Wainwright, K. Baker, and J. S. Mattick. 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* **19**:4008.
- García-Martínez, J., A. J. Martínez-Murcia, F. Rodríguez-Valera, and A. Zorraquino. 1996. Molecular evidence supporting the existence of two major groups in uropathogenic *Escherichia coli*. *FEMS Immunol. Med. Microbiol.* **14**:231–244.
- Gürtler, V., and H. D. Barrie. 1995. Typing of *Staphylococcus aureus* strains by PCR-amplification of variable-length 16S-23S rDNA spacer regions: char-

- acterization of spacer sequences. *Microbiology* **141**:1255–1265.
13. Gürtler, 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.
 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. Guttman, D. S., and D. E. Dykhuizen. 1994. Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. *Science* **266**:1380–1383.
 16. Harvey, S., C. W. Hill, C. Squires, and C. L. Squires. 1988. Loss of the spacer loop sequence from the *rnb* operon in the *Escherichia coli* K12 subline that bears the *relA1* mutation. *J. Bacteriol.* **170**:1235–1238.
 17. Jensen, M. A., J. A. Webster, and N. Straus. 1993. Rapid identification of bacteria on the basis of polymerase chain reaction-amplified ribosomal DNA spacer polymorphisms. *Appl. Environ. Microbiol.* **59**:945–952.
 18. Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, London, United Kingdom.
 19. Lawrence, J. G., H. Ochman, and D. L. Hartl. 1991. Molecular and evolutionary relationships among enteric bacteria. *J. Gen. Microbiol.* **137**:1911–1921.
 20. Leblond-Bourget, N., H. Philippe, I. Mangin, and B. Decaris. 1996. 16S rRNA and 16S to 23S internal transcribed spacer sequence analyses reveal inter- and intraspecific *Bifidobacterium* phylogeny. *Int. J. Syst. Bacteriol.* **46**:102–111.
 21. Lehner, A. F., S. Harvey, and C. W. Hill. 1984. Mapping and spacer identification of rRNA operons of *Salmonella typhimurium*. *J. Bacteriol.* **160**:682–686.
 22. Liu, S.-H., and K. E. Sanderson. 1996. Highly plastic chromosomal organization in *Salmonella typhi*. *Proc. Natl. Acad. Sci. USA* **93**:10303–10308.
 23. Martínez-Murcia, A. J., A. I. Antón, and F. Rodríguez-Valera. Concerted evolution by homogenization of the rRNA multigene family may induce sequence convergences responsible for chronometric distortions and conflicting phylogenetic trees. Submitted for publication.
 24. Milkman, R., and M. M. Bridges. 1993. Molecular evolution of the *Escherichia coli* chromosome. IV. Sequence comparisons. *Genetics* **133**:455–468.
 25. Nelson, K., T. S. Whittam, and R. K. Selander. 1991. Nucleotide polymorphism and evolution in the glyceraldehyde-3-phosphate dehydrogenase gene (*gapA*) in natural populations of *Salmonella* and *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **88**:6667–6671.
 26. Noller, H., and M. Nomura. 1996. Ribosomes, p. 167–186. 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, 2nd ed. ASM Press, Washington, D.C.
 27. Normand, P., C. Ponsonnet, X. Nesme, M. Neyra, and P. Simonet. 1996. ITS analysis of prokaryotes. p. 1–12. In A. D. L. Akkermans, J. D. van Elsas, and F. J. de Bruijn (ed.), *Molecular microbial ecology manual*. Kluwer Academic Publishers, Norwell, Mass.
 28. Ochman, H., and R. K. Selander. 1984. Standard reference strains of *Escherichia coli* from natural populations. *J. Bacteriol.* **157**:690–693.
 29. Pace, N. R., and A. B. Burgin. 1990. Processing and evolution of the rRNAs, p. 417–425. 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.
 30. Sechi, L. A., and L. Daneo-Moore. 1993. Characterization of intergenic spacers in two *rnm* operons of *Enterococcus hirae* ATCC 9790. *J. Bacteriol.* **175**:3213–3219.
 31. Selander, R. K., D. A. Cougant, and T. S. Whittam. 1987. Genetic structure and variation in natural populations of *Escherichia coli*, p. 1625–1648. In F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger, (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
 32. 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*. In F. G. Priest, A. Ramos-Cormenzana, and B. J. Tindall (ed.), *Bacterial diversity and systematics*. Plenum Press, New York, N.Y.
 33. Sharp, P. M. 1991. Determinants of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium*: codon usage, map position, and concerted evolution. *J. Mol. Evol.* **33**:23–33.
 34. Stolfus, A., J. F. Leslie, and R. Milkman. 1988. Molecular evolution of the *Escherichia coli* chromosome. I. Analysis of structure and natural variation in a previously uncharacterized region between *trp* and *tonB*. *Genetics* **120**:345–358.
 35. Thampapillai, G., R. Lan, and P. R. Reeves. 1994. Molecular evolution in the *gnd* locus of *Salmonella enterica*. *Mol. Biol. Evol.* **11**:813–828.
 36. Whiley, R. A., B. Duke, J. M. Hardie, and L. M. C. Hall. 1995. Heterogeneity among 16S-23S rRNA intergenic spacers of species within the '*Streptococcus milleri* group.' *Microbiology* **141**:1461–1467.