RNA (2002), *8*:1294–1307. Cambridge University Press. Printed in the USA.

DOI: 10.1017/S1355838202023014

The dispersal of five group II introns among natural populations of *Escherichia coli*

LIXIN DAI and STEVEN ZIMMERLY
Department of Biological Sciences, University of Calgary, Calgary, Alberta T2N 1N4, Canada

ABSTRACT

Group II introns are self-splicing RNAs that also act as retroelements in bacteria, mitochondria, and chloroplasts. Group II introns were identified in *Escherichia coli* in 1994, but have not been characterized since, and, instead, other bacterial group II introns have been studied for splicing and mobility properties. Despite their apparent intractability, at least five distinct group II introns exist naturally in *E. coli* strains. To illuminate their function and learn how the introns have dispersed in their natural host, we have investigated their distribution in the ECOR reference collection. Two introns were cloned and sequenced to complete their partial sequences. Unexpectedly, southern blots showed all ECOR strains to contain fragments and/or full-length copies of group II introns, with some strains containing up to 15 intron copies. One intron, *E.c.*I4, has two natural homing sites in IS*629* and IS*911* elements, and the intron can be present in one, both, or neither homing site in a given strain. Nearly all strains that contain full-length introns also contain unfilled homing sites, suggesting either that mobility is highly inefficient or that most full-length copies are nonfunctional. The data indicate independent mobility of the introns, as well as mobility via the host DNA elements, and overall, the pattern of intron distribution resembles that of IS elements.

Keywords: ECOR strains; retroelement; reverse transcriptase; ribozyme

INTRODUCTION

Group II introns are retroelements found in the genomes of bacteria, mitochondria, and chloroplasts (Bonen & Vogel, 2001; Belfort et al., 2002). Intron mobility occurs by site-specific insertion into defined target sites (retrohoming), or, at much lower frequencies, insertion into noncognate sites (retrotransposition). Group II intron retroelements consist of a self-splicing intron RNA structure, which encodes a multifunctional protein that facilitates intron splicing and intron mobility. The intron-encoded protein is a reverse transcriptase (RT) containing the seven domains common to all RTs, but also contains a domain X that aids in splicing of the intron in vivo, and a Zn domain, which provides a nuclease activity used in mobility. The mobility mechanism is well established and requires catalytic activities of both the self-splicing intron RNA and the intron-encoded protein. First, the intron reverse splices into the sense strand of the double-stranded DNA target site, and then the Zn domain cleaves the antisense strand and the RT reverse transcribes the intron using the cleaved DNA as a primer (Bonen & Vogel, 2001; Belfort et al., 2002).

Group II intron retroelements were discovered relatively recently in bacterial genomes. The first identification was achieved through PCR screening using degenerate primers (Ferat & Michel, 1993), which identified introns in a cyanobacterium (*Calothrix*) and a proteobacterium (*Azotobacter*). Because these bacteria are relatives of the ancestors of eukaryotic organelles, the finding suggested that mobile introns might have spread from bacteria to organelles. By now, due to the many bacterial genomes sequenced, it is clear that group II introns are widely dispersed throughout the eubacterial kingdom (Martinez-Abarca & Toro, 2000b; Dai & Zimmerly, 2002).

Bacterial group II introns differ in several respects from organellar introns. Unlike organellar introns, bacterial introns are not located in conserved genes, but are associated with mobile DNAs. The bacterial introns are often inserted outside of genes, and one subclass of introns inserts exclusively after transcriptional terminators rather than into ORFs. In contrast to organellar introns, all known bacterial group II introns encode RT ORFs and are either active retroelements or inactive

Reprint requests to: Dr. Steven Zimmerly, Department of Biological Sciences, University of Calgary, 2500 University Dr. NW, Calgary, Alberta T2N 1N4, Canada: e-mail: zimmerly@acs.ucalgary.ca.

derivatives. Finally, over half of bacterial group II intron sequences are fragments, suggesting that the introns survive by constantly inserting into new locations rather than assuming a relatively stable position in a conserved gene, as occurs in organelles. Together these observations led us to suggest that group II introns in bacteria are adapted to function mainly as retroelements (Dai & Zimmerly, 2002).

Group II introns were discovered in *Escherichia coli* in 1994, and were named IntA, IntB, IntC, and IntD (Ferat et al., 1994). Dot blot screening of strains of the ECOR collection indicated that only 18 of the 72 ECOR strains contain an intron. Full sequence was initially reported only for IntB, with partial sequences reported for IntA, IntC, and IntD (Ferat et al., 1994). Subsequently IntD was sequenced independently in a *Shigella* pathogenicity island (Rajakumar et al., 1997), and in *E. coli* plasmid pB171, where it is associated with a second homing site (Tobe et al., 1999). A fifth intron was identified in the virulence plasmid of *E. coli* 0157:H7 (Burland et al., 1998). A complete description of the five introns is presented in the Results section, where sequencing of IntA and IntC permits all five introns to be described together. Also, in this article, to be consistent with our previous names for group II introns across species (Dai & Zimmerly, 2002), we refer to the five *E. coli* introns as *E.c.*I1, *E.c.*I2, *E.c.*I3, *E.c.*I4, and *E.c.*I5 (instead of IntA, IntB, IntC, and IntD (Ferat et al., 1994) and "unnamed" (Burland et al., 1998)).

It is striking that although *E. coli* group II introns were among the first bacterial introns discovered, no further work has been reported, presumably because they were not tractable genetically or biochemically. Instead, the bacterial intron Ll.ltrB of *Lactococcus* has become the preferred system to study mobility and splicing (Bonen & Vogel, 2001; Belfort et al., 2002). In order to more fully understand the role and history of the five *E. coli* group II introns in their natural hosts, we have revisited the issue of their distribution among natural strains with a much more thorough analysis. The data indicate that group II introns are more prevalent than previously realized and have a highly heterogeneous and complex distribution, with up to 15 intron copies present in some strains. The introns appear to have spread among the population through a combination of independent mobility events and indirect mobility via their host DNAs. In addition to illuminating the spread of these introns, this study should facilitate their experimental characterization in *E. coli*.

RESULTS

Completion of sequencing of two introns and description of all five

Full sequences of *E.c.*I1 and *E.c.*I3 were obtained by PCR amplification of the introns based on predicted flanking exon sequences and host strains (Ferat et al., 1994), followed by cloning and sequencing (see Material and Methods). Based on the completed sequences, all five introns can be described. *E.c.*I1, *E.c.*I2, and *E.c.*I3 are related to each other and belong to the previously defined subclass "bacterial class D." (Intron subclasses are based on phylogenetic groupings of the intron-encoded ORFs, but each subclass also has a distinct intron secondary structure; Toor et al., 2001; Zimmerly et al., 2001.) Typical of bacterial class D, the introns are IIB-like in RNA structure (Fig. 1A, B), and the ORFs lack a Zn domain. Although the Zn domain is required for mobility of *Lactococcus* Ll.ltrB (Cousineau et al., 1998), many bacterial introns do not contain Zn domains and at least one of these is efficiently mobile nevertheless (Martinez-Abarca & Toro, 2000a). *E.c.*I1 and *E.c.*I2 are 67% identical in DNA sequence, whereas *E.c.*I3 is much less related, with 44% identity to *E.c.*I2 over its ORF. *E.c.*I1 sequence from ECOR43 has a stop codon in RT domain 6 and a frame shift in domain X, suggesting loss of mobility and splicing functions.

*E.c.*I1 and *E.c.*I2 are inserted into an ISEc1 element, which is itself contained within an Recombination hot spot (Rhs) element (Ferat et al., 1994). ISEc1 was previously called an H-repeat, but has been renamed ISEc1 because of its resemblance to IS elements (Mahillon & Chandler, 1998). The Rhs element is an ~6-kb DNA found in five copies in the sequenced K-12 genome. There are at least eight varieties of Rhs elements (A–H), whose core sequences are >70% identical, but which differ in organization and lengths of spacer elements (Zhao et al., 1993; Bachellier et al., 1996; Wang et al., 1998). At the 3′ terminus of most Rhs elements is an ISEc1 copy, which is flanked by 11 bp inverted repeats. *E.c.*I1 is inserted 4 bp after the upstream inverted repeat, whereas *E.c.*I2 is inserted near the 3′ end of the ISEc1 ORF. *E.c.*I3 is also located in an IS element, IS*679*.

*E.c.*I4 belongs to the subclass "bacterial class A," and its only other relative is the essentially identical intron in the closely related bacterium *Shigella* (99.4% identity). A possible secondary structure is shown in Figure 1C. According to information in the databases, the intron has two homing sites in IS*629* and IS*911* elements, which share only 60% identity (Fig. 2D). (We use the term "homing site" to denote the insertion site for all five introns in this study. Although homing has not been experimentally demonstrated, it is suggested because the introns are repeatedly found within the same flanking sequences.)

Finally, *E.c.*I5 belongs to "chloroplast-like class 1" and has a standard IIB1 intron structure (Fig. 1D). *E.c.*I5 is the only *E. coli* intron whose RT ORF encodes a Zn domain. Although the immediate relatives of *E.c.*I5 are bacterial, other members of the class reside mainly in chloroplasts and algal mitochondria, suggesting horizontal transfers (Zimmerly et al., 2001).

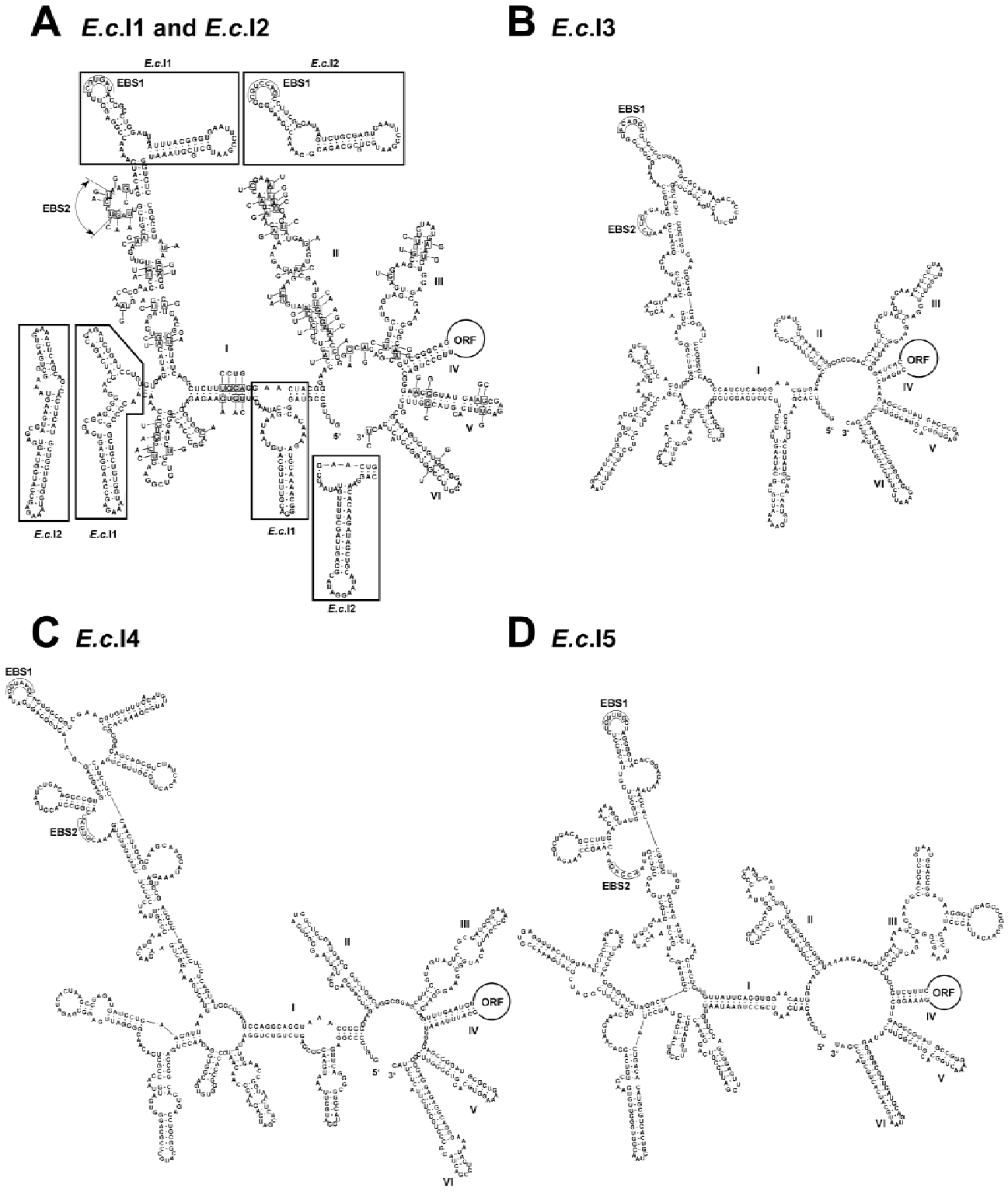

FIGURE 1. Secondary structure models for *E. coli* introns. RNA domains I to VI are shown for all structures, with the ORF looped out of domain IV. Exon binding sequence 1 (EBS1) and exon binding sequence 2 (EBS2) are indicated in domains I, and their complementary intron binding sequences 1 and 2 (IBS1, IBS2) are shown in exon sequences of Figure 2. See Toor et al. (2001) for detailed descriptions of intron structures and structural classes. **A:** Structural model for *E.c.*I1 with differences indicated for *E.c.*I2. The structure is IIB-like with some IIA and novel features. **B:** Structural model for *E.c.*I3, also typical of bacterial class D, but not highly conserved in sequence compared to *E.c.*I1 and *E.c.*I2. **C:** Structural model of *E.c.*I4, with characteristics of both IIA and IIB introns. The model is less reliable than other structures because there is no comparative information from related introns. **D:** Structural model for *E.c.*I5, with standard IIB1 features typical of chloroplast-like class 1.

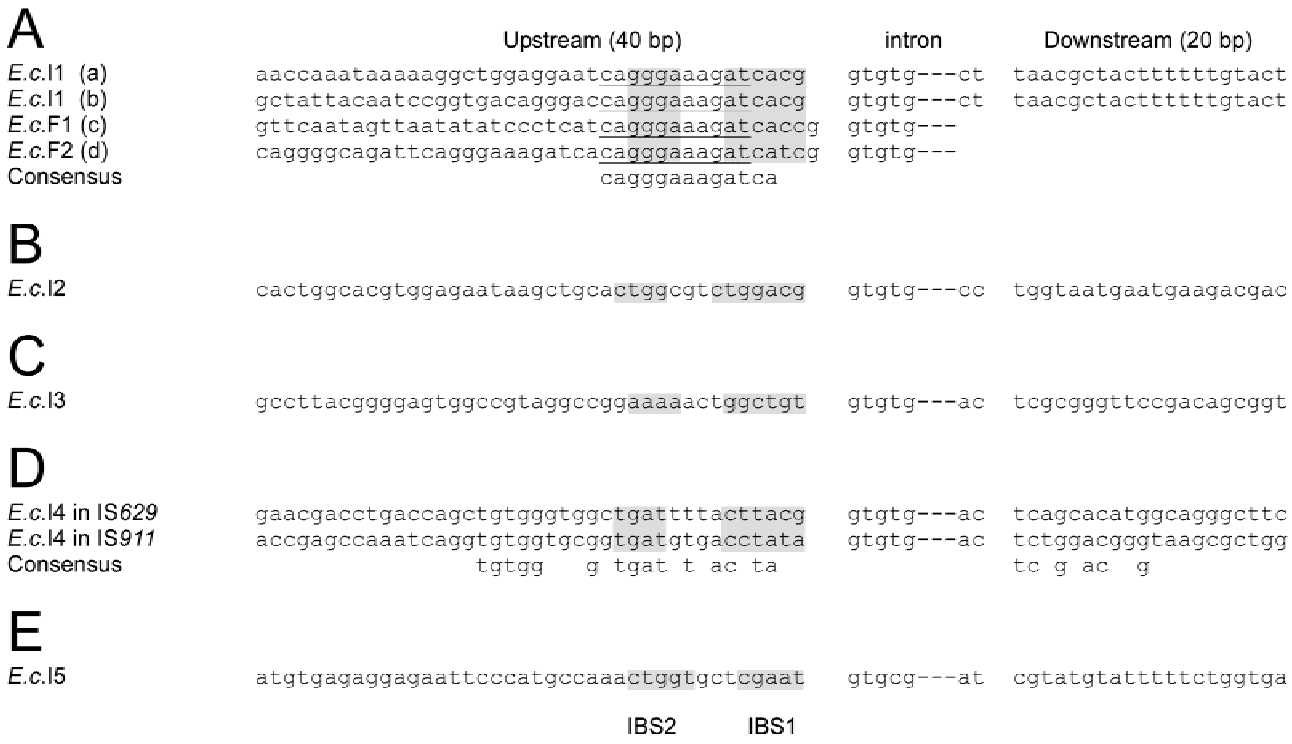

FIGURE 2. Flanking sequences of introns. Intron sequences are abbreviated as "gtgtg– – –ct" or a variation, and consensus sequences are shown when there is more than one flanking sequence. Intron binding sequences 1 and 2 (IBS1, IBS2) are indicated by shaded boxes, and pair with EBS1 and EBS2 sequences in Figure 1. **A:** Flanking sequences for two full-length *E.c.*I1 introns and two fragments. Sequence "a" is found for *E.c.*I1 in ECOR17, 22, and 67; sequence "b" is found in ECOR31 and ECOR43; sequences "c" and "d" are from intron fragments in ECOR11 (GenBank Accession No. AF044503) and ECOR45 (Accession No. AF044501), respectively. The 11-bp inverted repeat of ISEc1 is underlined. **B:** Flanking sequence of *E.c.*I2. **C:** Flanking sequence of *E.c.*I3. **D:** The two insertion sites for *E.c.*I4 in IS*629* and IS*911* elements. The target sites share 60% identity between the typical homing site boundaries of −25 and +10. **E:** Flanking sequence of *E.c.*I5.

Strategy for defining the intron content of ECOR strains

Intron content of the ECOR strains was inferred mainly through southern blots. Intron copy number was determined using intron probe and restriction enzymes that cut far from the intron (i.e., beyond the host exons), which is likely to yield a different size band for each intron copy (Fig. 3A). Conversely, homing sites were identified using restriction enzymes that cut near the intron (i.e., in the surrounding exons), which will collapse all intron copies into a single band if they are inserted into the known homing site. Bands of unexpected sizes are either introns in an unknown homing site, intron fragments, or introns in a polymorphic but known homing site. Quantitation of the bands in the two digests allows an estimation of copy number per band. Filled versus unfilled homing sites were then addressed by stripping the blots and reprobing with exon sequence (Fig. 3B). Unfilled homing sites correspond to a band of defined size, whereas intron in known homing sites would align with the signal in the previous hybridization (Fig. 3A). Other bands that align in the two panels are interpreted as being introns associated with a polymorphic, known homing site or an intron fragment associated with the known homing site. Finally, the presence and absence of predicted intron–exon and exon–exon junctions, as inferred by the blots, were confirmed by PCR reactions.

*E.c.*I4

*E.c.*I4 content was first determined for the six strains previously reported to carry the intron. Copy number was indicated by the *Eco*RI/*Pvu*II digestion, and it was concluded that there is one intron copy in ECOR10 and ECOR40, but unexpectedly, 8–15 copies in ECOR9, 38, 39, and 41 (Fig. 4A; Table 1). Homing sites for the copies were revealed by *Hin*dIII/*Pvu*II digests. For example, ECOR39 is estimated to contain one copy in the IS*911* homing site (1.9-kb band, square symbol) and approximately seven copies in the IS*629* homing site (2.0-kb band, round symbol). Interestingly, ECOR9, 38, 39, and 41 appear to have introns in both homing sites, whereas ECOR10 and ECOR40 have intron in only the IS*911* site (Table 1). (Data for ECOR40 and ECOR41 are based on PCR data and experiments be-

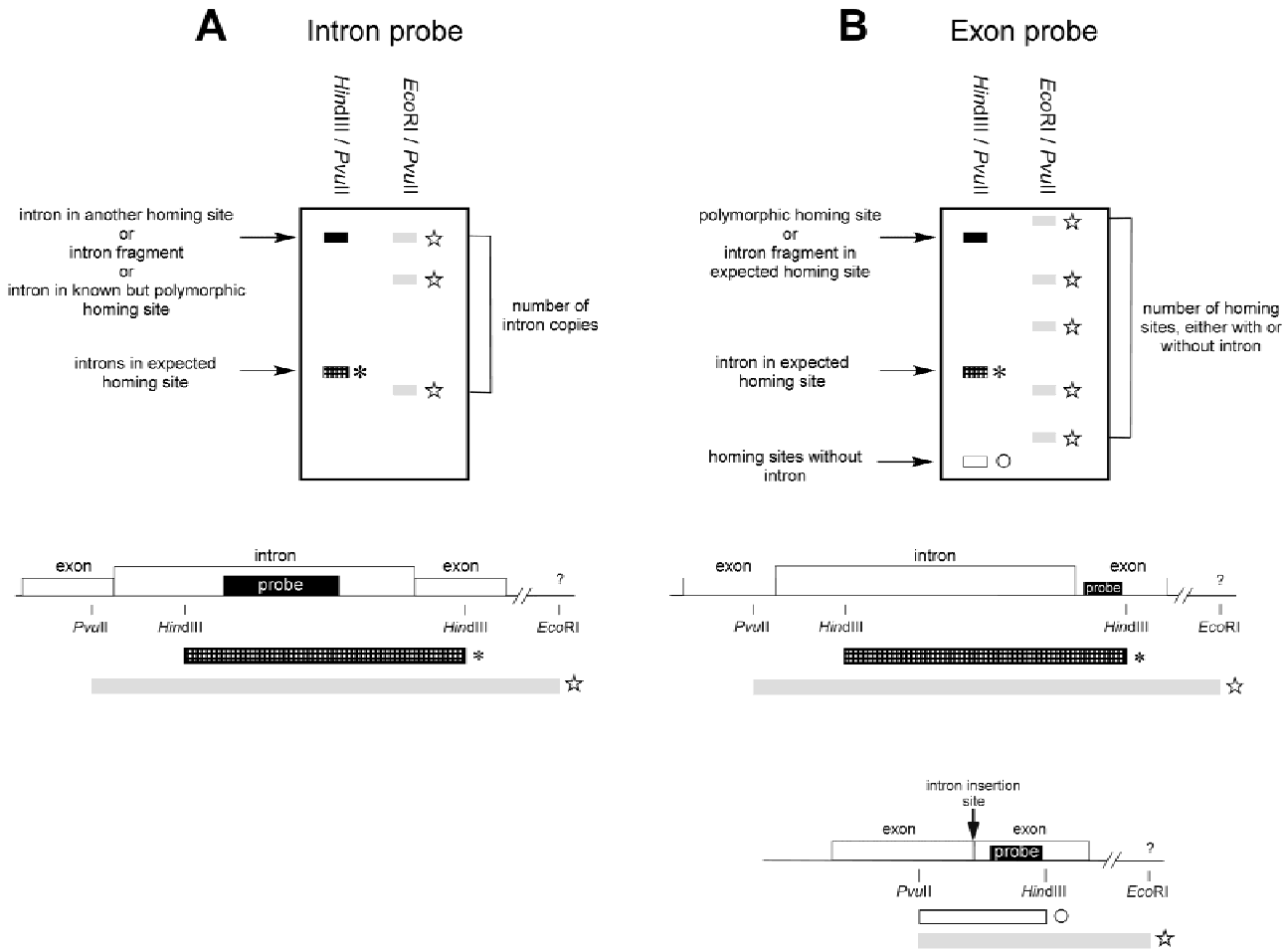

FIGURE 3. Southern blot strategy for determining intron copy number and homing site identity. **A:** Intron copy number is inferred from a southern blot using an intron probe and a restriction digest that cuts outside of the host exons. In this example, the *Eco*RI site is outside the host exon and is thus likely to produce a different band for each intron copy (star symbol). The homing site is determined using a restriction digest that cuts within the host exons, which will collapse all intron copies into a single band if they are located in the known homing site (asterisk symbol). Bands of other sizes are either intron copies in an unknown homing site, intron fragments associated with the known homing site, or introns in a known but polymorphic homing site (solid black bands). Quantitation of bands by phosphorimaging estimates copies represented by each band. **B:** The number of filled and unfilled homing sites is determined by hybridizing the same blot with exon probe. Unfilled homing sites are represented by a band of known size (open circle symbol), and introns in the known homing site are represented by the same band as in **A** (asterisk symbol). Bands of unexpected size that align in **A** and **B** are either intron fragments associated with the known homing site, or introns in a polymorphic but known homing site (solid black bands).

low in addition to the southern blots.) The presence and absence of intron–exon junctions for the two homing sites were confirmed by PCR for all six strains.

Bands other than 1.9 and 2.0 kb in the *Hin*dIII/*Pvu*II lanes might represent intron fragments, introns inserted into polymorphic IS*629* or IS*911* sites, or introns inserted into a third homing site. A potential third homing site suggested by a closely related intron fragment in *Yersinia pestis* (GenBank Accession No. AF074611) was screened for with PCR, but the potential intron–exon junction was not found (not shown). Additional unknown insertion sites were searched for using inverse PCR, but the only amplification product (ECOR38, *Eco*RV/*Pvu*II digest) was found upon sequencing to be a polymorphic IS*629* homing site. Based on additional data below, it is probable that all bands of unexpected sizes in the *Hin*dIII/*Pvu*II lanes are introns in polymorphic IS*629* or IS*911* elements, or intron fragments associated with IS*629* or IS*911* elements rather than introns in a third insertion site.

To determine whether intron-containing strains also contain unfilled homing sites, southern blots were stripped and reprobed with IS*629* exon sequence (Fig. 4B). Surprisingly, all strains except ECOR9 contain unfilled homing sites (456 bp) in numbers at least equal to intron-containing copies, based on hybridization strength (Fig. 4B; Table 1). Some bands other than 2.0 kb or 456 bp align with bands in Figure 4A (black round symbols), suggesting they are introns in polymorphic IS*629* sites or intron fragments associated with IS*629* sites. For ECOR40, a single hybridizing band of unexpected size was confirmed by PCR to be a full-

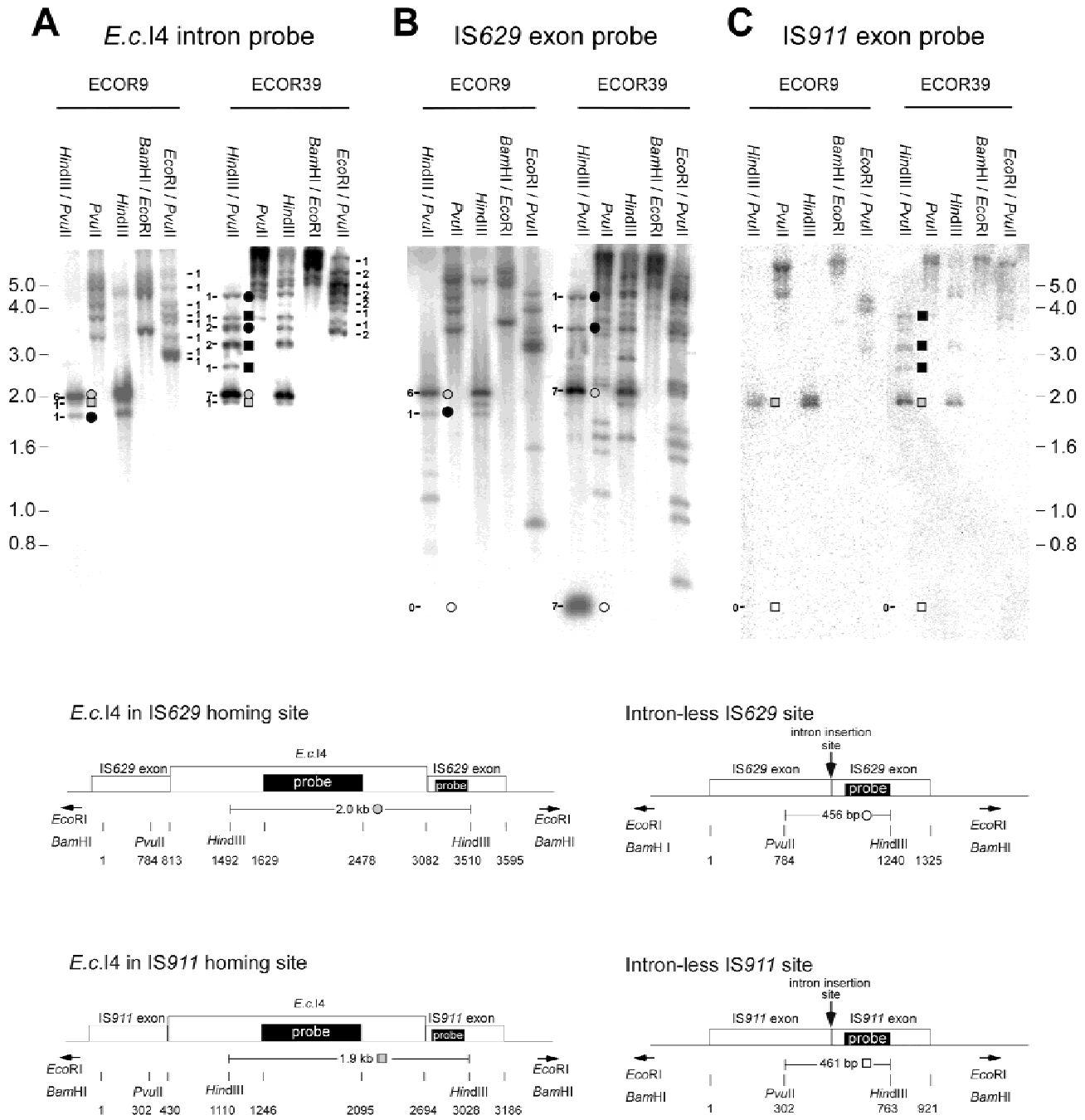

FIGURE 4. Representative southern blots inferring intron copy number and homing sites for *E.c.*I4. Restriction maps of filled and unfilled IS*629* and IS*911* homing sites are shown below (not drawn to scale). Complete data are summarized in Table 1. **A:** Intron copy number and identification of homing sites. Genomic DNAs of ECOR strains were digested with the indicated restriction enzymes and hybridized with an 849-bp *E.c.*I4 intron probe (black box in diagram below). Intron copy numbers are inferred from *Eco*RI/*Pvu*II digests combined with phosphorimaging quantitation of bands in the *Eco*RI/*Pvu*II and *Hin*dIII/*Pvu*II lanes (estimated copies per band are indicated in the two lanes). The *Hin*dIII/*Pvu*II digests identify introns in IS*629* homing sites (2.0-kb bands; gray circle symbols) and IS*911* homing sites (1.9-kb bands; gray square symbols). Thus, the example blots indicate that ECOR9 contains 8 intron copies, with 6 full-length copies in IS*629* and 1 in IS*911*, whereas ECOR39 contains 15 intron copies, with 7 full-length copies in IS*629* and 1 in IS*911*. **B:** Filled and unfilled IS*629* homing sites. Blots from **A** were stripped (except in this example, the ECOR9 blot is new) and probed with a 200-bp IS*629* exon probe (black box in diagram below). A 456-bp band (open circle symbol) indicates unfilled IS*629* sites, and a 2.0-kb band in the *Hin*dIII/*Pvu*II lane (gray circle symbol) indicates IS*629* sites filled with intron. Other bands aligning with **A** (black circle symbols) are inferred to be introns inserted into polymorphic IS*629* sites or fragments associated with IS*629*. Remaining bands are polymorphic IS*629* copies or IS*629* fragments. Copy number estimates of intron-less IS*629* sites were based on phosphorimaging quantitation of the 456-bp band relative to other bands in the *Hin*dIII/*Pvu*II digest. ECOR9 and ECOR39 are inferred to contain zero and seven unfilled IS*629* homing sites, respectively. **C:** Filled and unfilled IS*911* sites. Blots were stripped and reprobed with a 308-bp IS*911* exon probe. A 461-bp band in the *Hin*dIII/*Pvu*II lane (open square symbol) indicates unfilled IS*911* sites, and a 1.9-kb band (gray square symbol) indicates an IS*911* site containing intron. Additional bands aligning with **A** (black square symbols) are predicted to be intron fragments associated with IS*911* sites or introns in polymorphic IS*911* sites.

TABLE 1. Summary of intron and homing site content of ECOR strains.

	ECOR strain				
	17	22	31	43[a]	67
*E.c.*I1					
Total # *E.c.*I1 introns[b]	~2	~2	~2	~2	~4
Full-length introns inserted in ISEc1 homing site[c]	1	1	1	1	1
Probable intron fragments[d]	~1	~1	~1	~1	~3
Upstream sequence[e]	(a)	(a)	(b)	(b)	(a)

	ECOR strain										
	44[a]	47	9	10	17	20	22	24	31	43	67
*E.c.*I2											
Total # *E.c.*I2 introns[b]	1	1	0	0	0	0	0	0	0	0	0
Full-length introns inserted in ISEc1 homing site[c]	1	1	0	0	0	0	0	0	0	0	0
Probable intron fragments[d]	0	0	0	0	0	0	0	0	0	0	0
Intronless homing sites[f]	2	3	≥1	≥1	≥1	≥1	≥1	≥1	≥1	≥1	≥1

	ECOR strain									
	9	17	23	24	52[a]	60	10	20	38	39
*E.c.*I3										
Total # *E.c.*I3 introns[b]	~2	1	~2	~2	~2	~2	0	0	0	0
Full-length introns inserted in IS*679* homing site[c]	1[g]	0	1	1[g]	1	1	0	0	0	0
Intron fragments or introns in a polymorphic site[h]	~1	1[i]	~1	~1	~1	~1	N/A	0	N/A	N/A
Intronless homing sites[f]	1[g]	0	1[g]	1–4[g]	1[g]	1[g]	≥1	≥1	≥1	≥1

	ECOR strain												
	9	10	38	39	40	41	17	20	23	24	31	43	52
*E.c.*I4													
Total # *E.c.*I4 introns[b]	~8	1	~15	~15	~1	~10	0	0	0	0	0	0	0
Full-length intron inserted in IS*629*[c]	~6	0	~7	~7	0	~7	0	0	0	0	0	0	0
Intron fragments or introns in polymorphic IS*629* site[j]	~1[k]	0	~3	~3	~1[l]	~1	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Full-length intron inserted in IS*911*[c]	~1	1	~1	~1	0	0	0	0	0	0	0	0	0
Intron fragments or introns in polymorphic IS*911* site[m]	0	0	~4	~4	0	~2[n]	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Intronless IS*629* sites[f]	0	~4	~7	~7	~11	~3	0	~1	0	~8	0	0	0
Intronless IS*911* sites[f]	0	~1	0	0	~1	0	≥1	≥1	≥1	≥1	≥1	≥1	≥1

[a] The strain from which the full-length intron was sequenced.
[b] The total number of intron copies as determined by phosphorimaging quantitation of bands in southern blots. The copies are either full length or fragments.
[c] The number of introns inserted at the known homing site, as determined by southern blots. The copies are probably full length.
[d] Probable intron fragments, judging from weakly hybridizing bands of unexpected sizes in southern blots.
[e] Sequence upstream of the inverted repeat of ISEc1. "a" has no BLAST match in Genbank; "b" is 95% identical to RhsE (see Figure 2).
[f] The number of unfilled homing sites, as determined either by PCR (≥1) or by phosphorimaging quantitation of bands in southern blots (specific number).
[g] The complete homing site sequence is present; however, the upstream IS*679* exon is truncated.
[h] Bands of unexpected size in southern blots that are probably either intron fragments associated with a known homing site, or full-length introns in a known but polymorphic homing site.
[i] Probable internal fragment truncated at both 5′ and 3′ ends.
[j] Either an intron fragment associated with an IS*629* element or a full-length intron in a polymorphic IS*629* site.
[k] The short size of the band in the southern blot indicates an intron fragment rather than a polymorphic site.
[l] PCR amplification and sequencing of flanks show that the single copy of *E.c.*I4 in ECOR 40 is a full-length intron in a polymorphic IS*629* site.
[m] Either an intron fragment associated with an IS*911* element or an intron in a polymorphic IS*911* site.
[n] Intron is truncated at the 5′ end.

length *E.c.*I4 copy in a polymorphic IS*629* site. Sequencing of one junction showed five polymorphisms out of 70 bp in the upstream IS*629* exon, but no polymorphisms in the intron (not shown).

The finding that there are many unfilled homing sites is surprising because intron homing in yeast, *Lactococcus,* and *Sinorhizobium* is quite efficient (Moran et al., 1995; Mills et al., 1997; Martinez-Abarca & Toro, 2000a). One would expect that, over time, all potential homing sites should be filled. To confirm the sequences of the unfilled homing sites, the intronless sites were PCR amplified from ECOR10, 38, 39, 40, and 41, as well as ECOR20 and 24, which do not contain intron. At least one clone from each strain was sequenced for at least 40 bp on each side of the intron insertion site. All sequences were as expected from positions −35 to +25 relative to the insertion site, although there were polymorphisms outside of this region. Thus there are no polymorphisms that would affect intron insertion. The fact that homing sites remain unfilled suggests that the introns are mobile at extremely low rates, that they are regulated to transpose under rare circumstances, or possibly that most intron copies are completely immobile due to nonobvious mutations.

Hybridizations with IS*911* exon probe also showed unfilled IS*911* homing sites for two strains (Fig. 4C; Table 1). Again, unfilled homing sites were PCR amplified and sequenced to confirm the absence of mutations in the putative homing site. A summary of all information obtained for the *E.c.*I4 introns by southern blots and PCR experiments is presented in Table 1.

Finally, all 72 ECOR strains were tested for the presence of *E.c.*I4 by southern blots using a full-length intron probe that would detect all intron fragments. We reasoned that because some strains contain up to 15 intron copies, the original dot blot experiments might have overlooked some intron copies. In fact, all ECOR strains cross-hybridized with the full-length probe, although the six previously identified strains cross-hybridized most strongly. To confirm that the weak signals in most of the 72 strains are due to intron fragments or single intron copies, selected strains were screened by PCR to test for the presence of the first 794 bp of the intron (5′ intron terminus), or the last 299 bp (3′ intron terminus). Only 1 out of 11 strains tested contained the 5′ intron terminus, whereas 8 of 11 contained the 3′ intron terminus, strongly suggesting that the weak hybridization signals are fragments. Southern blots of three strains showed that ECOR37, which contained both 5′ and 3′ termini according to PCR, does indeed contain a full-length intron copy, whereas ECOR14 and ECOR56, which contain only 3′ termini by PCR, showed different signals of unpredicted sizes, and are likely nonidentical 3′ fragments. Consistent with the apparent prevalence of *E.c.*I4 fragments, the sequenced *E. coli* 0157:H7 genome contains a 76-bp fragment of the 5′ terminus of *E.c.*I4, whereas the sequenced K12 genome contains a 1,758 bp fragment truncated at its 3′ end by an IS*5* copy. Together, there are at least four forms of *E.c.*I4 fragments in *E. coli* strains. In summary, *E.c.*I4 is much more widely distributed in *E. coli* strains than realized, with all strains containing at least a fragment of the intron, and a small proportion containing up to 15 copies of full-length intron in two natural homing sites.

*E.c.*I1

Five ECOR strains were previously reported to contain *E.c.*I1 (ECOR17, 22, 31, 43, and 67; Ferat et al., 1994). Based on southern hybridizations (not shown), each strain contains one full-length intron in the expected ISEc1 homing site. In addition, each strain probably contains at least one fragment, as suggested by weak hybridizations of unexpected size (Table 1). Interestingly, the intron insertion site is 15 bp from the 5′ end of ISEc1, suggesting an abbreviated homing site extending to only −15 rather than the typical −25. To investigate the boundary of the homing site, inverse PCR was used to obtain sequence of the upstream flank. Two different 5′ sequences were obtained. ECOR17, 22, and 67 have an upstream sequence “a,” which has no match in GenBank, and ECOR31 and ECOR43 have the sequence “b,” which matches RhsE sequence (Fig. 2A). Two more fragments of *E.c.*I1 were found in GenBank entries for ECOR11 and ECOR45, each with different upstream sequences (Fig. 2A). Together, the four upstream sequences suggest no homing site conservation beyond −15, although it remains possible that the ISEc1 transposed after intron insertion. All 72 ECOR strains were screened by southern blots with a full-length *E.c.*I1 probe, but because the *E.c.*I1 and *E.c.*I2 probes cross-hybridized, they will be discussed together below.

*E.c.*I2

Southern blots showed *E.c.*I2 to have a simple distribution in ECOR44 and ECOR47 (not shown), with each strain containing one full-length intron copy in the predicted ISEc1 homing site, and two (ECOR44) or three (ECOR47) intronless homing sites. Intronless sites were sequenced and confirmed not to contain mutations preventing intron insertion. Screening of the 72 ECOR strains with full-length *E.c.*I1 and *E.c.*I2 probes showed very similar signals due to cross-hybridization (67% identity overall, with scattered regions of >90% identity). DNA from 57 of the 72 strains cross-hybridized with both probes (see Fig. 5). PCR screening of eight weakly hybridizing strains showed that all contain the 5′ terminus of *E.c.*I1 but not the 3′ terminus, whereas none contain the 5′ or 3′ terminus of *E.c.*I2. Therefore, we judge that most or all of the weak signals represent fragments of *E.c.*I1 rather than *E.c.*I2. Consistent with this

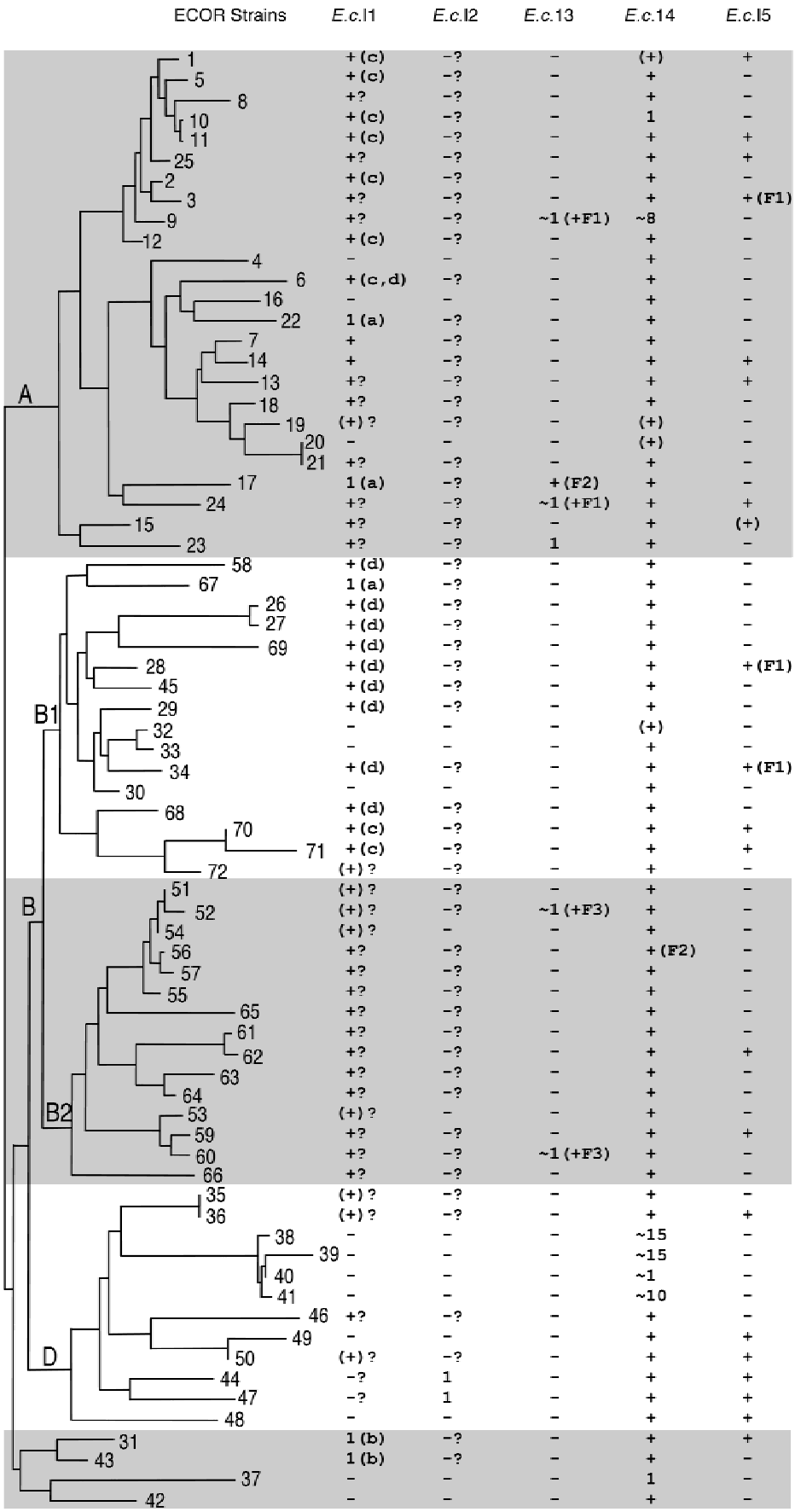

FIGURE 5. *See caption on facing page.*

conclusion, there are no fragments of *E.c.*I2 in the databases, but there are 5′ fragments of *E.c.*I1 in the completely sequenced *E. coli* 0157:H7 genome, and in GenBank entries for ECOR11 (Accession No. AF044503) and ECOR45 (Accession No. AF044501). *E.c.*I1 fragments in *E. coli* 0157:H7 and ECOR45 are identical and are variant ISEc1 copies, including an upstream 5′ inverted repeat plus 193 bp of *E.c.*I1 (Fig. 2A), followed by an intact ISEc1. The 5′ fragment in ECOR11 is similar in organization but has a different upstream sequence (Fig. 2A) and only 159 bp of *E.c.*I1 before the intact ISEc1. It is possible that the *E.c.*I1 fragments are carried by the host Rhs and/or ISEc1 elements, and because it has been proposed that Rhs elements propagated vertically after being introduced into the *E. coli* population (Hill et al., 1995), this might account for the hybridization in most strains. Vertical inheritance of the fragments was supported by PCR screening for ECOR11-type or ECOR45-type fragments ("c" and "d" in Figure 2A), which showed fragment "c" to be present in nearly all strains screened from subgroup A (Fig. 5), whereas fragment "d" is present in nearly all strains screened from subgroup B1, and neither fragment is present in 11 strains tested from subgroups B2 and D. A small degree of horizontal transfer of fragments might have occurred also, because there is one "d" fragment in subgroup A (ECOR6) and two "c" fragments in subgroup B1 (ECOR70 and 71). However, the *E.c.*I1 fragments in subgroups A and B1 did not originate from full-length introns in the same subgroups because they have distinct upstream flanks. This suggests that they are propagating independently from the full-length copies, possibly by ISEc1 movement or recombination events.

*E.c.*I3

Southern hybridizations showed a single copy of *E.c.*I3 in the expected homing site in ECOR9, 23, 24, 52, and 60, and a second copy in ECOR9, 24, 52, and 60 that is either an intron fragment or an intron in a polymorphic IS*679* element. ECOR17 showed a single weakly hybridizing band that is probably an intron fragment, which was supported by the failure to amplify either the 5′ or 3′ terminus of the intron (not shown). Unfilled homing sites were identified by PCR and southern blots, and all five strains that contain full-length intron were found to contain unfilled homing sites (Table 1). Interestingly, although the unfilled homing sites in ECOR9, 23, 24, 52, and 60 are intact in the region of intron insertion, the upstream IS*679* exon is truncated, suggesting that the intron might have inserted independently into the same truncated IS element. Southern blots of all 72 ECOR strains did not identify additional copies of *E.c.*I3.

*E.c.*I5

*E.c.*I5 is present as a single copy in the virulence plasmid p0157 of *E. coli* 0157:H7 (Burland et al., 1998). Southern blot screening showed *E.c.*I5 to be present in 19 of the 72 ECOR strains (not shown), but they appear to be mainly fragmented forms. Of 10 hybridizing strains tested, only 3 contain the 5′ terminus (756 bp) according to PCR, and none contain the 3′ terminus (192 bp), suggesting that nearly all hybridizations are due to internal fragments rather than full-length copies. Southern blots with the three strains containing the 5′ intron terminus (ECOR3, 28, 34) indicated identical fragments in all three strains, and PCR verified the expected upstream homing site sequence. It is possible that a small set of fragmented forms of *E.c.*I5 has spread through the population via host mobile DNAs.

Inference of intron spread based on phylogenetic relationships of the ECOR strains

The ECOR collection represents natural diversity of *E. coli* strains (Ochman & Selander, 1984), and has been used to examine distribution and spread of a number of elements. Based on inferred relationships among the ECOR strains, Rhs elements were concluded to be inherited mainly vertically (Hill et al., 1995), whereas retrons appear to have invaded *E. coli* several times by sporadic horizontal transfers (Herzer et al., 1990). IS*1*, IS*2*, IS*3*, IS*4*, IS*5*, and IS*30* have highly variable content among closely related strains, and their distribution has been explained by occasional horizontal transfers, probably by plasmid vectors, followed by multiple transposition events to amplify copy numbers (Green et al., 1984; Sawyer et al., 1987; Boyd & Hartl, 1997).

Group II intron distribution in ECOR strains is diagramed in Figure 5 according to phylogenetic relation-

FIGURE 5. Phylogenetic distribution of the five group II introns in ECOR strains. The phylogenetic tree was redrawn from Herzer et al. (1990) and is based on neighbor-joining analysis of multilocus enzyme electrophoresis data. Major groupings A, B1, B2, and D are shaded. Numbers indicate intron copy number based on southern blots similar to Figure 4. + and − indicate the presence or absence of intron based on southern blot hybridizations with full-length intron probe (not shown). For *E.c.*I1 and *E.c.*I2, + and − indicate that intron presence was confirmed by PCR, and question marks indicate the probable source of hybridization as described in the text. Parentheses indicate weak hybridization. (a), (b), (c), and (d) in the *E.c.*I1 column refer to upstream flanking sequence (Figure 2). F1, F2, and F3 in *E.c.*I3 and *E.c.*I4, and *E.c.*I5 columns represent distinguishable forms of fragments, as judged by southern blots.

ships inferred by Herzer et al. (1990). Four basic groupings were specified based on multilocus enzyme electrophoresis data, with variable electrophoretic mobilities of enzymes used as phylogenetic markers. An independent estimate of strain phylogeny using the same method is mainly consistent with these groupings (Selander et al., 1987), although the differences do not affect any conclusions discussed here.

Figure 5 suggests examples of both vertical and horizontal inheritance for group II introns. Vertical inheritance is suggested by related strains that contain the same introns: *E.c.*I1 in ECOR31 and ECOR43; *E.c.*I2 in ECOR44 and ECOR47; and *E.c.*I4 in ECOR38, 39, 40, and 41. In the ECOR38/39/40/41 group, copy number was apparently amplified after the initial introduction. The fragment forms "b," "c," and "d" of *E.c.*I1 also provide an example of vertical inheritance, probably via the host Rhs elements with occasional losses.

Examples of horizontal transfers are suggested by sporadic distribution of introns in different ECOR subclasses: *E.c.*I1 in ECOR17, 22, 67, and 31/43; *E.c.*I3 in ECOR9, 23, 24, 52, and 60; *E.c.*I4 in ECOR9, 10, 37, and 38/39/40/41. Horizontal transfer of specific fragments is also likely for some *E.c.*I5 fragments. Although it is not possible to exclude the spotty distribution as being due to vertical inheritance with frequent intron loss, vertical inheritance alone seems unlikely in light of known horizontal transfers of IS elements that are the intron hosts. Horizontal transfers via host mobile DNAs also allows for specific forms of intron fragments to be propagated through a population (e.g., *E.c.*I5 fragments).

DISCUSSION

We have shown that group II introns are more prevalent in *E. coli* populations than previously realized, that they have highly variable composition in related strains, and that they are very frequently fragmented. We propose that all five group II introns are active retroelements in *E. coli* despite their apparent experimental intractability, and that their distribution is due to a combination of horizontal and vertical inheritance for both full-length and fragmented forms of introns.

Comparison with RmInt1 distribution in *Sinorhizobium meliloti*

RmInt1 of *Sinorhizobium meliloti* is the only other group II intron to be investigated within a population (Martinez-Abarca et al., 1998; Muñoz et al., 2001). Like *E. coli*, RmInt1 is variable in copy number, with 0–11 copies in different strains; however, unlike *E. coli* introns, RmInt1 is quite prevalent, with most *S. meliloti* strains containing at least one full-length RmInt1 copy. Interestingly, many intron-containing *S. meliloti* strains contain host IS elements unfilled with intron, as deduced from hybridizations (Martinez-Abarca et al., 1998). Although the hybridizations were not determined to represent unmutated homing sites, the combined observations in *S. meliloti* and *E. coli* suggest that it may be common for intron-containing bacteria to contain unfilled homing sites.

Evidence for independent mobility of the introns

Because all five *E. coli* introns are located in mobile DNAs, it has been speculated that their dispersal may rely on host element mobility rather than independent mobility (Ferat et al., 1994). Several observations in this study argue for independent mobility, the clearest being *E.c.*I4 composition in the ECOR38/39/40/41 grouping, in which each strain has a different intron composition. Importantly, there are three distinguishable homing sites: an IS*911* site, a IS*629* site in ECOR9, 38, 39, and 41 (2.0-kb band in Fig. 4A), and a polymorphic IS*629* site in ECOR40. Therefore, *E.c.*I4 has inserted independently into new homing sites at least twice within the ECOR38/39/40/41 grouping. The *E.c.*I4 composition in ECOR9 also supports independent mobility, because the intron is located in both IS*629* and IS*911* sites, whereas no related ECOR strains contain the intron. Similarly, *E.c.*I1 appears to have independently inserted into four different ISEc1 homing sites (Fig. 2A), although it is also possible that the ISEc1 element transposed after intron insertion. For *E.c.*I2, *E.c.*I3, and *E.c.*I5, there is no information about exon polymorphisms, and therefore no direct evidence for independent mobility; however, their distribution suggests that they may also be independently mobile at some frequency.

How are intron fragments formed and propagated?

Intron fragmentation may occur by several mechanisms. Based on analysis of group II intron sequences in GenBank (Dai & Zimmerly, 2002), one common mechanism event is truncation through the insertion of other mobile elements. For example, the *E.c.*I4 fragment in the K12 genome is truncated at its 3′ terminus by the insertion of an IS*5* element. Other group II intron fragments are flanked by stem-loops, which might be forms of mobile DNA themselves, or might be formed during the truncation event. Alternatively, fragments could be formed during intron insertion, with some 5′ truncated fragments being due to incomplete reverse transcription, analogous to 5′ truncations of non-LTR retroelements.

The prevalence of the intron fragments might be explained in two ways. First, each fragment might represent an independent insertion and fragmentation event.

In this case, the large number of intron fragments would imply very frequent rates of mobility and fragmentation. Alternately, intron fragmentation events might be comparatively rare, and specific subsets of intron fragments might propagate vertically with occasional loss, or horizontally via a host mobile DNA. The detection of identical fragments of *E.c.*I1 and *E.c.*I5 in multiple, sometimes distantly related strains supports the second possibility. On the other hand, there are at least three versions of *E.c.*I1 fragments and four versions of *E.c.*I4 fragments, and so it is clear that the introns have been fragmented multiple times.

Why are there so many unfilled homing sites?

The most perplexing finding in the study is that almost all strains that contain full-length intron also contain unfilled homing sites. Group II introns in other bacteria including *Lactococcus* (LI.ltrB) and *Sinorhizobium* (RmInt1) are mobile in vivo at efficiencies of roughly 10–30% (Mills et al., 1997; Martinez-Abarca & Toro, 2000a) and it would seem that, over time, all homing sites in *E. coli* should be filled. The abundance of unfilled sites suggests either that the introns are mobile at extremely low rates, or that most intron copies are nonfunctional because of unrecognized mutations. A similar unexplained observation has been noted for mobile group I introns, which also have not saturated their homing sites in natural populations (Edgell et al., 2000).

One possibility is that many group II introns might be inactivated by mutations, as deleterious mutations are quite common among group II introns in many bacterial species (Dai & Zimmerly, 2002), and one of the five *E. coli* introns clearly contains mutations in one strain (*E.c.*I1). Still, the other four sequenced copies appear functional, and it is questionable whether mutations are present in all cases ($>$50 full-length intron copies are present in strains with unfilled homing sites). A simpler explanation is that the introns are intrinsically mobile at extremely low rates (i.e., lower than IS element mobility); however, such low mobility would not explain why some strains contain up to 15 copies of intron. Amplification of introns via host IS elements is a possible explanation, but IS elements generally do not transpose frequently (Green et al., 1984). A third possibility is that the introns might be highly mobile during short periods followed by latency, as has been reported for the *Sinorhizobium* intron RmInt1. Introduction of homing site into intron-containing *Sinorhizobium* strains results in homing efficiencies of $\sim$10% within 24 h, but there is no further increase (Martinez-Abarca & Toro, 2000a). Possible conditions in *E. coli* that might regulate group II intron mobility include new introduction of either homing site or intron DNA into a cell, or environmental conditions such as starvation or stress.

Overall model for spread of the introns through the *E. coli* population

Overall, we propose that group II introns have moved through populations of *E. coli* by a combination of sporadic horizontal transfers mediated by IS elements and plasmids, and vertical inheritance with amplification through homing. Because the five *E. coli* introns are not closely related to each other, with the exception of *E.c.*I1 and *E.c.*I2, the introns must have been introduced several times into *E. coli* from other bacterial or organellar sources (Zimmerly et al., 2001). After arriving in *E. coli*, the introns increased in copy number in a given lineage by independent homing into unfilled sites, or possibly by transposition of the host mobile DNAs. Recombination and gene conversion are also possible mechanisms for intron amplification or rearrangement. Rhs elements, for example, are suggested to be reshuffled through recombination (Zhao et al., 1993). The group II introns in *E. coli* also appear to have been exchanged horizontally among related strains and species, with both full-length and fragmented introns spreading via host mobile DNA elements. Judging from the number of unfilled homing sites, the introns clearly are not mobile at high efficiencies, as can sometimes occur in other bacteria, but it remains possible that the introns might be mobile under specific conditions, or that a small fraction of intron copies might be highly mobile. In this regard, *E.c.*I4 appears most promising for experimental characterization because it is the most prevalent, being found in all ECOR strains, and in one strain, ECOR9, all homing sites are filled by intron.

MATERIALS AND METHODS

Growth of *E. coli* and preparation of genomic DNA

The ECOR collection was obtained from Dr. Ken Sanderson (University of Calgary). *E. coli* strains were grown under standard conditions in LB without antibiotics. For preparations of genomic DNA, cells were grown to stationary phase, resuspended in TE (10 mM Tris, pH 8.0, 1 mM EDTA), and lysozyme was added to a final concentration of 1%. Cells were incubated at 37 °C for 15 min, followed by addition of sarkosyl to a 1% final concentration, and proteinase K to 50 mg/mL. The lysate was incubated at 60 ° for 3 h, and then the DNA was extracted by phenol, and precipitated by 0.6 vol isopropanol. Samples were resuspended in TE and treated with 0.1 mg/mL RNase A and 0.05 mg/mL RNase T1 at 37 ° for 1 h, followed by phenol extraction, precipitation with isopropanol and 0.3 M NaOAc, and resuspension in H_2O. For some ECOR strains, 5 mM EDTA were included during cell lysis stages to inhibit endogenous nucleases, which allowed much better recovery of DNA.

PCR and DNA sequencing

PCR was with either *Taq* (Invitrogen, Carlsbad, California) or *Pfu* DNA polymerases (Stratagene, La Jolla, California) ac-

cording to the manufacturer's protocols. Inverse PCR was according to Ausubel et al. (1999) with digestion of genomic DNA with *Dra*I, *Pvu*II, and *Eco*RI. Sequencing was done manually by standard methods, or by the DNA sequencing center of the University of Calgary. Primers for inverse PCR for *E.c.*I1 were: 5′-TGTCGGGGAGGTTCCGGTGGCCTACTC, 5′-GCCTGCCACCATCTCTTCACAAGCTTATGC. Primers to amplify homing sites were: *E.c.*I2: 5′-CGGGATCCCCAG AGCATGACGTTACGCAATGAGTG, 5′-CCCTGCCTTGAAT ACCTTATCATTCGTC; *E.c.*I3: 5′-CGGACCGGAAAGGTAT CCATCCACAAAATC, 5′-GAGGTCAGATCAACTTTCCAGG GCAACAG; *E.c.*I3 (short): 5′-AGTGGCCGTAGGCCGGAAA AACTGGCTGT, 5′-GAGGTCAGATCAACTTTCCAGGGCA ACAG; *E.c.*I4 (IS*629*): 5′-GCGTCAGTTGTTACGGGAAG GAATCAGG, 5′-CGCCTTATAGGCCAGTGACACATACTG; and *E.c.*I4 (IS*911*): 5′-TCAGGTGTGGTGCGGTGATGTGA CCTATA, 5′-TAACCCACCACCGGCATCCATTCGTTCTTC. Primers to amplify 5′ termini of introns were: *E.c.*I1: 5′-GTG TGCCGTAGGCACAAAGATGCAAAACG, 5′-GCCTGCCAC CATCTCTTCACAAGCTTATGC; *E.c.*I2: 5′-GTGTGCCGCA GGAACACAAGATAGCTGC, 5′-ACTTCCTGAAGAAAGTC GGTTCCAG; *E.c.*I4: 5′-GTGTGCCCGGAGTTCAGGGCG GGCAT, 5′-TTTCGCCGCTCTTTGATGAAAGCGTGATCC GCGCC; and *E.c.*I5: 5′-GTGCGACATGAAGTCGCCTGA ATAA, 5′-TGCTCCCTCCACCTTGCCATGTTTACCTTCTC GTG. Primers to amplify 3′ termini of introns were: *E.c.*I1: 5′-CGGAAATTCCCTGGTGTCGTTATGCAGATG, 5′-AGTG AGTAGGCCACCGGAACCTCC; *E.c.*I2: 5′-TCAGCCGATAC AAGACAAGAGCCTC, 5′-GGTGAGTAGGCCGCAGGAACC TCCC; *E.c.*I4: 5′-CCGAAACCGGGTCAGAGCAAAACGTG, 5′-GTAGAGTAAGAGGCAGGGCGTAGT; and *E.c.*I5: 5′-GG AGGGAGCAACTGCCTCTCCAATTTG, 5′-ATCGGGTAGC AGTGATGCATTACTGCAT. Primers to amplify the 5′ junction of *E.c.*I4 (IS*629*) were: 5′-GCGTCAGTTGTTACGGGAAGG AATCAGG, 5′-CATCTCGGCTTAGTGATCTCGCCTCAATC. Primers to amplify the 3′ junction of *E.c.*I4 (IS*911*) were: 5′-GCGGCGAAATACTGTACCGGTTGGTG, 5′-TAACCCAC CACCGGCATCCATTCGTTCTTC. Primers to amplify internal fragments of *E.c.*I3 were: 5′-CTGTCCGATGAAGCATC GAAATCTTAG and 5′-ATACTGTTGGGATCCCCAGAATGC GCTCTT; 5′-CCATGCTGATTCATATGGTTACCGACCAGG and 5′-CCTCTCAAACGGTACCCCCCGATATTCTCTTT.

Southern blots

Southern blot experiments were done by standard methods according to Sambrook et al. (1989) with 50% formamide and at 42 °C. Probes for southern blots were made by PCR amplification of ECOR genomic DNA in 25 cycles in reaction volumes of 25 μL with 10 ng genomic DNA, 50 ng of each primer, 1 mM $MgCl_2$, 2.5 μL 2 mM dNTPs, 50 Ci [α-^{32}P]-dCTP (3,000 Ci/mmol; Amersham Biosciences, Piscataway, New Jersey) and *Taq* DNA polymerase (Invitrogen, Carlsbad, California) according to the supplier's protocol. Resulting probes had specific activities of approximately 100,000 cpm/mg DNA. Probes were purified from a 1% agarose gel.

Oligonucleotides used to make probes for the five introns are as follows. Partial intron sequence probes: *E.c.*I1: 5′-CGGAAATTCCCTGGTGTCGTTATGCAGATG, 5′-CCCGT CAGGCAAACGCGCTTTCCGGTCC; *E.c.*I2: 5′-GAGCCAC ATTTTCTGGCAGACTCTTACG, 5′-GGCAAACAACTGTG GACTTCGCTTCGC; *E.c.*I3: 5′-CCATGCTGATTCATATGGT TACCGACCAGG, 5′-CCTCTCAAACGGTACCCCCCGATAT TCTCTTT; and *E.c.*I4: 5′-ATGGCGTGAACAAAACAATGCT ACAGGC, 5′-ATCCTGTGCCCCAGAAAGATAAAGCCGTC. Full-length intron probes: *E.c.*I1: 5′-GTGTGCCGTAGGC ACAAAGATGCAAAACG, 5′-AGTGAGTAGGCCACCGGAA CCTCC; *E.c.*I2: 5′-GTGTGCCGCAGGAACACAAGATAGC TGC, 5′-GGTGAGTAGGCCGCAGGAACCTCCC; *E.c.*I3: 5′-GTGTGTCACGAAGGAGTACTTATGGC, 5′-GTCGAGTC AGCCGGAGAGAATTTCACT; *E.c.*I4: 5′-GTGTGCCCGGA GTTCAGGGCGGGCAT, 5′-GTAGAGTAAGAGGCAGGGCGT AGT; and *E.c.*I5: 5′-GTGCGACATGAAGTCGCCTGAATAA, 5′-ATCGGGTAGCAGTGATGCATTACTGCAT. Exon probes: *E.c.*I2: 5′-CGGGATCCCCAGAGCATGACGTTACGCAATG AGTG, 5′-CGGGATCCTCCAGACGCCAGTGCAGCTTATT CTCCA; *E.c.*I3: 5′-CGGACCGGAAAGGTATCCATCCACA AAATC, 5′-ACAGCCAGTTTTTCCGGCCTACGGCCACT; *E.c.*I4 (IS*629*): 5′-TCAGCACATGGCAGGGCTTCGT, 5′-CG CCTTATAGGCCAGTGACACATACTG; and *E.c.*I4, (IS*911*): 5′-CGGGATCCTACCTCGCCGTTGTTCTCGACCTGTT, 5′-TAACCCACCACCGGCATCCATTCGTTCTTC.

Bands in southern blots were quantitated by phosphorimaging (MacBAS, Fuji Medical Systems, Inc., Stamford, Connecticut). For *E.c.*I4, copy numbers were estimated by comparing quantitation of bands in *Hin*dIII/*Pvu*II and *Eco*RI/*Pvu*II lanes, with strong bands considered to represent multiple copies, and weak bands representing single copies. In cases where the *Hin*dIII/*Pvu*II and *Eco*RI/*Pvu*II lanes estimated different numbers of total introns, the copy numbers per band were adjusted as minimally as possible to result in equal estimations for the two lanes. For introns other than *E.c.*I4, there was no evidence for multiple full-length intron copies, and so weakly hybridizing bands of unexpected sizes were considered probable fragments.

Cloning and sequencing of *E.c.*I1 and *E.c.*I3

Full length copies of *E.c.*I1 and *E.c.*I3 with flanks were PCR amplified from genomic DNA of ECOR43 and ECOR52, respectively, with *Pfu* DNA polymerase. Amplification of *E.c.*I1 was with the primers 5′-CAGGGAAAGATCACGGTG and 5′-CTCTACTTTCCAGGCTTATCTGTAATCGGG, with an annealing temperature of 48 °C. Amplification of *E.c.*I3 was with the primers 5′-CGGACCGGAAAGGTATCCATCCACAAAATC and 5′-GAGGTCAGATCAACTTTCCAGGGCAACAG with an annealing temperature of 55 °C. Products were cloned and both strands were sequenced. Sequence data are deposited in GenBank (Accession Nos. AF512502 and AF512503). Both sequences are completely identical to partial sequences reported in Ferat et al. (1994), and the *E.c.*I3 sequence is 98.4% identical to an intron copy in *Shigella* (GenBank Accession No. AF348706), except that the *Shigella* intron is truncated by 34 bp at its 5′ end.

ACKNOWLEDGMENTS

We thank K. Sanderson, D. Taylor, and T. Tobe for gifts of strains and plasmids. This work was supported by the Canadian Institutes of Health Research (CIHR) and by the Alberta Heritage Foundation for Medical Research (AHFMR).

Received May 30, 2002; returned for revision
July 3, 2002; revised manuscript received July 30, 2002

REFERENCES

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K. 1999. *Current protocols in molecular biology.* New York: John Wiley & Sons, Inc.

Bachellier S, Gilson E, Hofnung M, Hill CW. 1996. Repeated sequences. In: Neidhardt FC, ed. *Escherichia coli and Salmonella: Cellular and molecular biology,* 2nd ed. Washington, DC: ASM Press. pp 2012–2040.

Belfort M, Derbyshire V, Parker MM, Cousineau B, Lambowitz AM. 2002. Mobile introns: Pathways and proteins. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, eds. *Mobile DNAII*. Washington, DC: ASM Press. pp 761–783.

Bonen L, Vogel J. 2001. The ins and outs of group II introns. *Trends Genet 17*:311–331.

Boyd EF, Hartl DL 1997. Nonrandom location of IS1 elements in the genomes of natural isolates of *Escherichia coli*. *Mol Biol Evol 14*:725–732.

Burland V, Shao Y, Perna NT, Plunkett G, Sofia HJ, Blattner FR. 1998. The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* 0157:H7. *Nucleic Acids Res 26*:4196–4204.

Cousineau B, Smith D, Lawrence-Cavanagh S, Mueller JE, Yang J, Mills D, Manias D, Dunny G, Lambowitz AM, Belfort M. 1998. Retrohoming of a bacterial group II intron: Mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell 94*:451–462.

Dai L, Zimmerly S. 2002. Compilation and analysis of group II intron insertions in bacterial genomes: Evidence for retroelement behavior. *Nucleic Acids Res 30*:1091–1102.

Edgell DR, Belfort M, Shub DA. 2000. Barriers to intron promiscuity in bacteria. *J Bacteriol 182*:5281–5289.

Ferat J-L, Le Gouar M, Michel F. 1994. Multiple group II self-splicing introns in mobile DNA from *Escherichia coli*. *CR Acad Sci Paris 317*:141–148.

Ferat J-L, Michel F. 1993. Group II self-splicing introns in bacteria. *Nature 364*:358–361.

Green L, Miller RD, Dykhuizen DE, Hartl DL. 1984. Distribution of DNA insertion element IS5 in natural isolates of *Escherichia coli*. *Proc Natl Acad Sci USA 81*:4500–4504.

Herzer PJ, Inouye S, Inouye M, Whittam TS. 1990. Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. *J Bacteriol 172*: 6175–6181.

Hill CW, Feulner G, Brody MS, Zhao S, Sadosky AB, Sandt CH. 1995. Correlation of Rhs elements with *Escherichia coli* population structure. *Genetics 141*:15–24.

Mahillon J, Chandler M. 1998. Insertion sequences. *Microbiol Mol Biol Rev 62*:725–774.

Martinez-Abarca F, Toro N. 2000a. Homing of a bacterial group II intron with an intron-encoded protein lacking a recognizable endonuclease domain. *Mol Microbiol 35*:1405–1412.

Martinez-Abarca F, Toro N. 2000b. Group II introns in the bacterial world. *Mol Microbiol 38*:917–926.

Martinez-Abarca F, Zekri S, Toro N. 1998. Characterization and splicing in vivo of a *Sinorhizobium meliloti* group II intron associated with particular insertion sequences of the IS*630*-Tc1/IS*3* retroposon superfamily. *Mol Microbiol 28*:1295–1306.

Mills DA, Manias DA, McKay LL, Dunny, GM. 1997. Homing of a group II intron from *Lactococcus lactis* subsp. lactis ML3. *J Bacteriol 179*:6107–6111.

Moran JV, Zimmerly S, Eskes R, Kennell JC, Lambowitz AM, Butow R, Perlman PS. 1995. Mobile group II introns of yeast mitochondrial DNA are novel site-specific retroelements. *Mol Cell Biol 15*:2828–2838.

Muñoz E, Villadas PJ, Toro N. 2001. Ectopic transposition of a group II intron in natural bacterial populations. *Mol Microbiol 41*:645–652.

Ochman H, Selander RK. 1984. Standard reference strains of *Escherichia coli* from natural populations. *J Bacteriol 157*:690–693.

Rajakumar K, Sasakawa C, Adler B. 1997. Use of a novel approach, termed island probing, identifies the *Shigella flexneri* she pathogenicity island which encodes a homolog of the immunoglobulin A protease-like family of proteins. *Infect Immun 65*:4606–4614.

Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: A laboratory manual*, 2nd ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Sawyer SA, Dykhuizen DE, DuBose RF, Green L, Mutangadura-Mhlanga T, Wolczyk DF, Hartl DL. 1987. Distribution and abundance of insertion sequences among natural isolates of *Escherichia coli*. *Genetics 115*:51–63.

Selander RK, Caugant DA, Whittam TS. 1987. Genetic structure and variation in natural populations of *Escherichia coli*. In: Neidhardt FC, ed. *Escherichia coli and Salmonella typhimurium: Cellular and molecular biology*. Washington DC: ASM Press. pp 1625–1648.

Tobe T, Hayashi T, Han CG, Schoolnik GK, Ohtsubo E, Sasakawa C. 1999. Complete DNA sequence and structural analysis of the enteropathogenic *Escherichia coli* adherence factor plasmid. *Infect Immun 67*:5455–5462.

Toor N, Hausner G, Zimmerly S. 2001. Co-evolution of the group II intron RNA structure with its intron-encoded reverse transcriptase. *RNA 7*:1142–1152.

Wang YD, Zhao S, Hill CW. 1998. Rhs elements comprise three subfamilies which diverged prior to acquisition by *Escherichia coli*. *J Bacteriol 180*:4102–4110.

Zhao S, Sandt CH, Feulner G, Vlazny DA, Gray JA, Hill CW. 1993. Rhs elements of *Escherichia coli* K-12: Complex composites of shared and unique components that have different evolutionary histories. *J Bacteriol 175*:2799–2808.

Zimmerly S, Hausner G, Wu X-c. 2001. Phylogenetic relationships among group II intron ORFs. *Nucleic Acids Res 29*:1238–1250.