JOURNAL OF BACTERIOLOGY, Nov. 1989, p. 6197–6205
0021-9193/89/116197-09$02.00/0

Vol. 171, No. 11

Characterization of the Cryptic Lambdoid Prophage DLP12 of *Escherichia coli* and Overlap of the DLP12 Integrase Gene with the tRNA Gene *argU*

DAVID F. LINDSEY, DAVID A. MULLIN,† AND JAMES R. WALKER*

Department of Microbiology, University of Texas, Austin, Texas 78712

Received 8 May 1989/Accepted 31 July 1989

The *argU* (*dnaY*) gene of *Escherichia coli* is located, in clockwise orientation, at 577.5 kilobases (kb) on the chromosome physical map. There was a cryptic prophage spanning the 2 kb immediately downstream of *argU* that consisted of sequences similar to the phage P22 *int* gene, a portion of the P22 *xis* gene, and portions of the *exo*, *P*, and *ren* genes of bacteriophage λ. This cryptic prophage was designated DLP12, for defective lambdoid prophage at 12 min. Immediately clockwise of DLP12 was the IS*3* $\alpha_4\beta_4$ insertion element. The *argU* and DLP12 *int* genes overlapped at their 3′ ends, and *argU* contained sequence homologous to a portion of the phage P22 *attP* site. Additional homologies to lambdoid phages were found in the 25 kb clockwise of *argU*. These included the cryptic prophage *qsr′* (P. J. Highton, Y. Chang, W. R. Marcotte, Jr., and C. A. Schnaitman, J. Bacteriol. 162:256–262, 1985), a sequence homologous to a portion of λ *orf*-194, and an *attR* homolog. Inasmuch as the DLP12 *att int xis exo P/ren* region, the *qsr′* region, and homologs of *orf-194* and *attR* were arranged in the same order and orientation as the lambdoid prophage counterparts, we propose that the designation DLP12 be applied to all these sequences. This organization of the DLP12 sequences and the presence of the *argU*/DLP12 *int* pair in several *E. coli* strains and closely related species suggest that DLP12 might be an ancestral lambdoid prophage. Moreover, the presence of similar sequences at the junctions of DLP12 segments and their phage counterparts suggests that a common mechanism could have transferred these DLP12 segments to more recent phages.

The *Escherichia coli argU* gene encodes a minor arginine tRNA, $tRNA^{Arg}_{UCU}$, which corresponds to the rare codon AGA and possibly AGG (18, 39, 51). (This gene was originally called *dnaY* because a temperature-sensitive mutation caused preferential inhibition of DNA synthesis at high temperature [24]; it is now redesignated *argU* after the convention of Fournier and Ozeki [17].) *argU* expression has several unusual characteristics. First, it is transcribed into a monocistronic message, whereas most tRNA genes are parts of operons (17). Second, transcription in vitro is terminated by a rho-dependent process to generate 180- and 190-nucleotide products. The precursors are presumably processed at the 5′ and 3′ ends to generate the 77-nucleotide mature form. Third, $tRNA^{Arg}_{UCU}$ is a minor species, suggesting inefficient transcription. However, the *argU* promoter should be strong, based on its sequence and spacing (23). This suggests negative control of *argU*, and a potential regulator site of 31 base pairs (bp) of dyad symmetry is located downstream of the promoter −10 sequence. Fourth, a 387-codon open reading frame (ORF) extends toward and overlaps the 3′ end of *argU* (51). This ORF is especially interesting because it contains the extremely high number of 13 AGA and AGG codons—the *argU* tRNA cognates. Understanding *argU* regulation and the relationship of the downstream ORF might provide clues about the role of *argU* in replication, which at present is unknown.

In this article, we demonstrate that the ORF adjacent to *argU* is a homolog of the phage P22 integrase gene. It is part of a cryptic prophage which is designated DLP12, which probably integrated into *argU*. The *argU*/DLP12 *int* pair are found in several *E. coli* strains and related species, suggesting that DLP12 might be an ancestral lambdoid phage.

MATERIALS AND METHODS

Bacterial strains, plasmids, and bacteriophages. *E. coli* strains B, C, and F and *Salmonella typhimurium* LT2 DB7000 were obtained from I. J. Molineux. *Enterobacter aerogenes* SD1, *Erwinia carotavora* EC153 *trp-9*, *Klebsiella pneumoniae* M5a1, and *Vibrio harveyi* B392 were obtained from J. W. Zyskind. Genomic DNA from a clinical *Shigella boydii* isolate (strain O-1392, serotype 5; Texas Department of Health) was provided by S. M. Payne.

Strain JM103 was the host for plasmids and M13 phages, and the *recBC sbcB* strain JC7623 (32) was used in the construction of an insertion mutant. Plasmid pSP6/T7-19 and bacteriophages M13mp8 and −9 were obtained from Bethesda Research Laboratories. Plasmid pUC4K (50) was from Pharmacia. Plasmid pDM1 was constructed by cloning a 4.5-kilobase (kb) *Hin*dIII *argU*$^+$ (*dnaY*$^+$) fragment from pLC22-8 into pBR322 (39). A 2.2-kb *Hin*cII-*Hin*dIII *argU*$^+$ fragment was subcloned from pDM1 into pUC9 cut with *Hin*cII and *Hin*dIII to generate pDC1 (Fig. 1A).

Recombinant DNA technology. Standard recombinant DNA techniques were used for the isolation of plasmid DNA, restriction endonuclease and *Bal*31 exonuclease digestions, ligations, transformations, and gel electrophoresis (37).

DNA sequencing. The nucleotide sequence was determined by the dideoxy chain termination method (46) according to the manual provided by Bethesda Research Laboratories. *Bal*31 deletions of the 2.2-kb *argU* fragment were constructed by opening plasmid pDC1 with *Bam*HI, treating with *Bal*31 for various time intervals, and cutting with *Hin*dIII to release a set of overlapping fragments which

* Corresponding author.
† Present address: Biology Department, Tulane University, New Orleans, LA 70118.

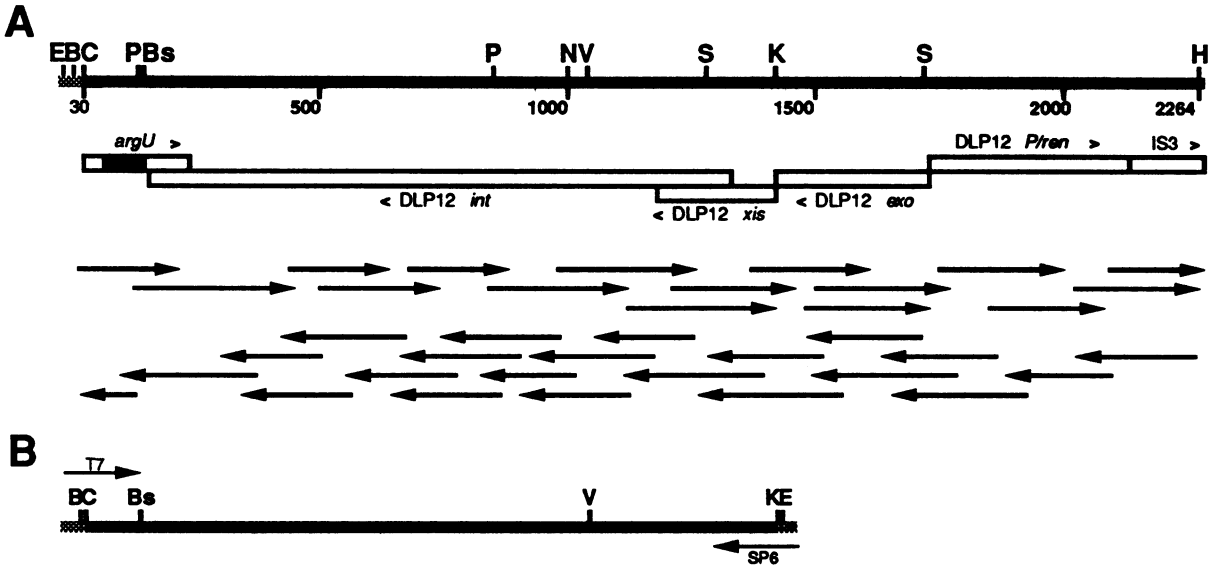


FIG. 1. (A) Organization and sequencing strategy of the *E. coli argU*/DLP12 region. The thick solid line represents chromosomal DNA present in pDC1; the stippled line represents a portion of the pUC9 polylinker region. Nucleotides are numbered according to GenBank Release 58. The sequence from positions 1 through 357 was presented in Garcia et al. (18) as −47 through 310. Open bars indicate positions of genes or regions; those labeled above the bars are oriented left to right (>), and those labeled below the bars have right-to-left orientation (<). The darkened position of the *argU* bar is the portion which corresponds to the mature tRNA. The direction and extent of the sequencing reactions are indicated by the arrows. Restriction enzymes: B, *Bam*HI; Bs, *Bss*HII; E, *Eco*RI; V, *Eco*RV; C, *Hin*cII; H, *Hin*dIII; K, *Kpn*I; N, *Nsi*I; P, *Pst*I; S, *Sma*I. (B) Plasmid pDC403, used in the construction of RNA probes, consists of the *Bam*HI-*Kpn*I fragment cloned into pSP6/T7-19. The direction of transcription from the T7 and SP6 promoters is indicated by arrows.

were cloned into M13mp8 cut with *Sma*I and *Hin*dIII. For fragments of the opposite strand, pDC1 was opened with *Hin*dIII, treated with *Bal*31, and cut with *Eco*RI, and the overlapping fragments were cloned into M13mp9 cut with *Hin*cII and *Eco*RI. The nucleotide sequence of the 2.2-kb *Hin*cII-*Hin*dIII genomic region of pDC1 was determined from these M13 derivatives according to the scheme shown in Fig. 1A.

Preparation of RNA probes. Radiolabeled RNA probes were prepared as described before (38). The 1.4-kb *Bam*HI-*Kpn*I fragment from pDC1 (Fig. 1B) was subcloned into pSP6/T7-19 to generate pDC403, which was linearized with appropriate enzymes and transcribed by using SP6 or T7 RNA polymerase in the presence of [α-^{32}P]CTP (800 Ci/mmol).

Isolation of genomic DNA and Southern analysis. Cells from 150-ml cultures were lysed and treated with RNase A and protease (5), except that proteinase K (200 μg/ml) was used and the treatment was continued overnight. The lysis mixture was extracted once with phenol, once with phenol-chloroform-isoamyl alcohol (25:24:1, vol/vol), and once with chloroform. The DNA was precipitated with ethanol, dissolved in TE (10 mM Tris [pH 8.0], 1 mM EDTA), and dialyzed against TE overnight.

*Hin*dIII-restricted genomic DNA (5 μg) was fractionated on a 0.7% agarose gel and subsequently transferred to GeneScreen (Dupont-NEN Products) according to the manufacturer's protocol. The membrane was prehybridized in 50% formamide–6× SSC (1× SSC is 0.15 M NaCl, 0.015 M sodium citrate)–1% sodium dodecyl sulfate (SDS)–5× Denhardt solution (1× Denhardt solution is 0.02% bovine serum albumin, 0.02% polyvinylpyrrolidone, 0.02% Ficoll)–250 μg of denatured salmon sperm DNA per ml at 50°C for 30 h. The membrane was then hybridized to 10^6 cpm of RNA probe per ml in fresh hybridization buffer at 50°C for 30 h. The membrane was subsequently washed twice in 1× SSC–0.1% SDS for 30 min each at room temperature and twice in 0.1× SSC–0.1% SDS for 30 min each at 65°C. The membrane was autoradiographed with Kodak XRP-1 film.

Isolation of RNA and Northern (RNA blot) analysis. RNA was isolated by a modification of the method of Bovre et al. (7). Cells were lysed in TES (10 mM Tris [pH 7.5], 10 mM EDTA, 100 mM NaCl)–0.5% SDS and treated with proteinase K (200 μg/ml) for 20 min at 42°C. Total cellular RNA was recovered by extraction with hot phenol, phenol-chloroform-isoamyl alcohol, and chloroform and then precipitated with ethanol. The RNA samples were then treated with RNase-free DNase I (100 U of RNA per ml) for 15 min at 37°C, reextracted with phenol-chloroform-isoamyl alcohol and then chloroform, and ethanol precipitated.

For Northern analysis, RNA was fractionated on 1.5% agarose–2.2 M formaldehyde gels, transferred to GeneScreen membranes according to the manufacturer's protocol, and hybridized to radiolabeled RNA probes transcribed from an appropriate template with T7 RNA polymerase as described before (38).

RESULTS

Nucleotide sequence and analysis of the *argU*/DLP12 region. The nucleotide sequence of both strands of the chromosomal insert in pDC1 was determined (Fig. 2). The sequence within 2 kb downstream of the *argU* gene contained regions similar to three different lambdoid phage DNA segments and at least a portion of an insertion element. DLP12 (for defective lambdoid prophage at min 12) will be used when designating the phage homologs found in this region.

Immediately downstream of *argU* there was an open reading frame (ORF) of 387 codons. The deduced amino acid sequence was 71% identical to the 387-amino-acid phage P22 integrase (*int* gene product) (Fig. 3A) and contained a 40-amino-acid sequence near the C-terminal end which is

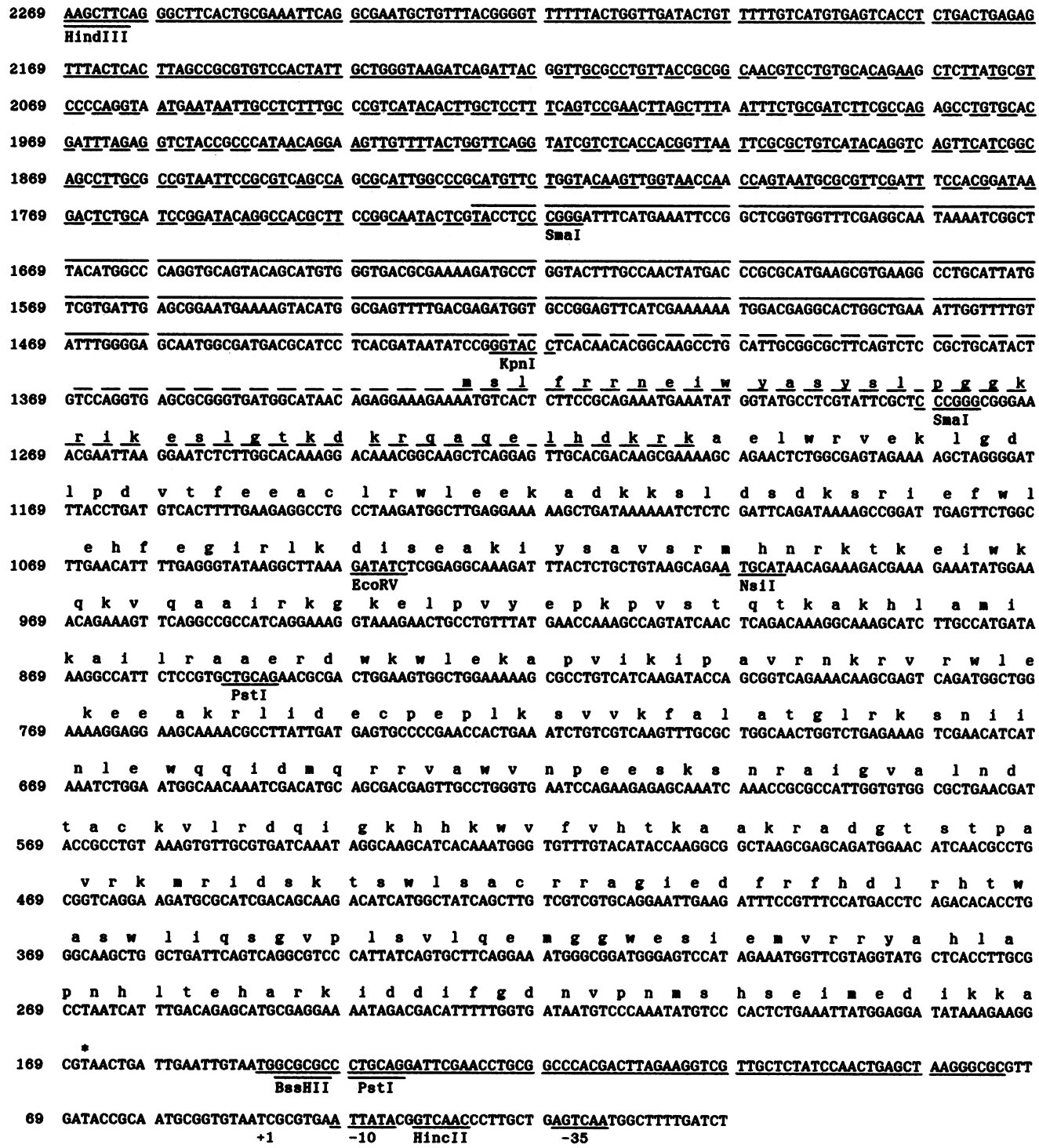
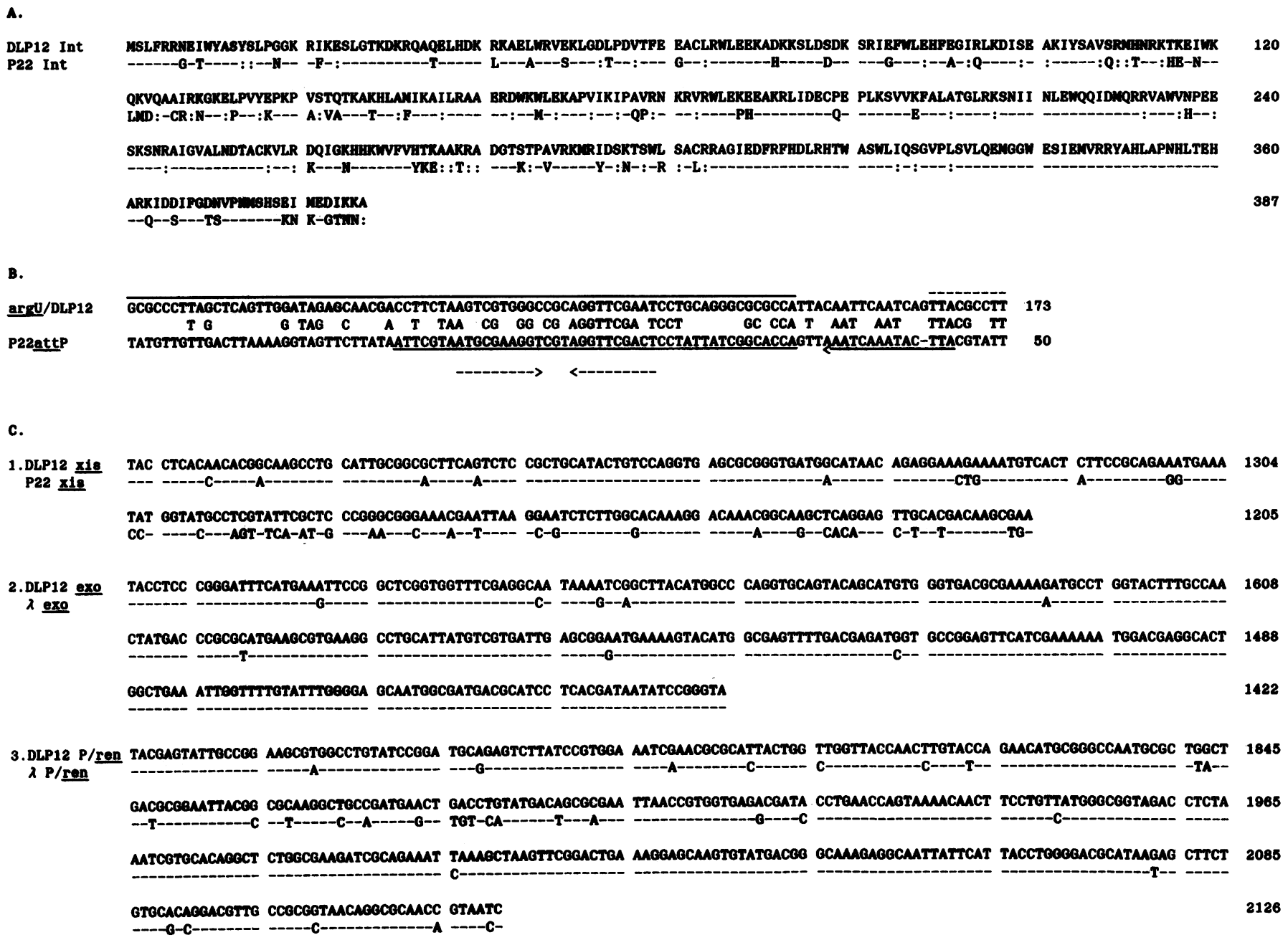


FIG. 2. Nucleotide sequence of the *argU*/DLP12 region and deduced amino acid sequence of DLP12 *int*. The DNA strand with the same sequence as the *int* mRNA is presented. Gene segments oriented 5′→3′ are designated by overlines; those oriented opposite to the strand presented are indicated by underlines. IS*3*, solid underline; DLP12 *P/ren*, dashed underline; DLP12 *exo*, solid overline; DLP12 *xis*, dashed overline. DLP12 *int* is shown translated. The DNA segment encoding mature *argU* tRNA sequences is indicated by a solid underline. Also shown are the *argU* promoter (−10, −35) and transcription start site (+1). Restriction sites shown in Fig. 1 are underlined and identified. These sequence data have been deposited in GenBank under accession number M27155.

highly conserved among the lambdoid phage integrases (1, 34). This gene presumably encodes an integrase and was designated DLP12 *int*. The *argU* and DLP12 *int* genes were arranged in opposite orientation, with the DLP12 *int* ORF terminating only 18 bp from the *argU* sequence encoding the 3′ CCA of the mature tRNA (Fig. 1, 2, and 3B). As summarized by Leong et al. (33), lambdoid phage attachment sites are adjacent to their *int* genes. Analysis of the


A.
DLP12 Int
P22 Int
B.
argU/DLP12
P22attP
C.
1.DLP12 xis
P22 xis
2.DLP12 exo
λ exo
3.DLP12 P/ren
λ P/ren

sequence downstream of the DLP12 *int* gene also demonstrated a region overlapping *argU* which was similar to approximately one-half of the P22 *attP* region (Fig. 3B). The P22 *attP* region consists of a core, containing two potential Int recognition sequences arranged as inverted repeats, flanked by two potential integration host factor (IHF)-binding sites (33) (Fig. 3B). The DLP12 presumptive *attP* half-site consisted of one region similar to the IHF-binding site consensus sequence plus roughly one-half of the core containing one potential Int core recognition sequence. Nucleotides which corresponded to $tRNA^{Arg}_{UCU}$ positions 51 through 77 constituted the DLP12 *attP* half-site.

Nucleotide sequence similarity to the phage P22 genome extended another 95 bp upstream of the DLP12 *int* gene. Positions 1329 through 1423 of DLP12 were 91.6% identical to the corresponding region of the P22 genome. This included a portion of the P22 excisionase and was designated DLP12 *xis*. In phage P22, the coding region for the Xis polypeptide is 351 bp long; the 3′ end of *xis* overlaps the 5′ end of *int* by 124 bp (34). DLP12 *xis*, including its 124-bp overlap with DLP12 *int*, consisted of 219 bp homologous to the 3′ end of P22 *xis* (Fig. 3C). DLP12 *xis* was contained within an "ORF" of 87 codons, of which an internal block of 74 codons were 70% homologous to amino acids 43 to 116 of P22 *xis*. However, this ORF did not contain an initiation codon and, because it encoded only a portion of Xis, might not encode a functional excisionase. On the other hand, DLP12 *xis* may contain sequences involved in the regulation of DLP12 *int* transcription. It is interesting that the 95 bp upstream of the DLP12 *int* ORF differed only 8.4% from the sequence of the homologous region of the phage P22 genome, while the *int* ORF sequences differed by 32%.

Two additional DLP12 segments were found upstream of *xis*. First, the sequence from position 1727 through 1422 was found to be 97.4% identical to the phage λ sequence (12) from positions 31627 through 31322 (Fig. 3C). In phage λ, this region encodes the C-terminal 92 residues of the 226-amino-acid exonuclease, the product of the *exo* gene, and 26 bp downstream of the *exo* ORF. Therefore, this region was designated DLP12 *exo*. Second, nucleotides from positions 1726 through 2126 in the DLP12 sequence were found to be 92.0% identical to nucleotides 39972 through 40372 in the phage λ sequence (Fig. 3C). In λ, this region encodes the C-terminal 102 residues of the 233-amino-acid *P* gene product and the N-terminal 31 residues of the 96-amino-acid *ren* gene product. This segment of DLP12 was therefore designated *P/ren*. Redfield and Campbell (42) have previously reported that a region counterclockwise of the *qsr′* defective prophage is homologous to λ *exo* and *P* region probes.

This region of DLP12 consisted of three contiguous segments, half-*att–int–xis*, *exo*, and *P/ren*, arranged in the same order and orientation as their phage P22 and λ counterparts, and resembled the left side of λ prophage with *attL*, *int*, *xis*, *exo*, and *P/ren* in left to right sequence. An analysis of the junctions between adjacent DLP12 segments (DLP12 *xis/exo* and DLP12 *exo/P*) and their phage counterparts revealed these interesting similarities (Fig. 4). First, the sequence within the P22 *xis* gene where homology with DLP12 *xis* ended was GAGGTA.AGGGGATTTCGGTT, where the period indicates the exact position where homology ended. Similarly, the sequence within the λ *exo* gene where homology with DLP12 *exo* ended was the very similar GAGG TA.AACGGGCATTTCAGTT. These two sequences follow the general pattern of GAGGTA.A($N_{0,2}$)GGGNATTTC NGTT, where ($N_{0,2}$) represents 0 or 2 nucleotides. Second, the λ *exo* and *P/ren* sequences which were homologous with DLP12 *exo* at the *xis-exo* junction and with DLP12 *P/ren* at the *exo-P/ren* junction, respectively, started with TAC. Third, the boundaries between the DLP12 *xis* and *exo* and between *exo* and *P/ren* repeated the 7-bp sequence GAGG-TAC. These similarities suggest that if DLP12 served as a primordial source of lambdoid information, a common mechanism served to transfer these fragments to more recent phages.

The IS*3* $\alpha_4\beta_4$ insertion element is adjacent to DLP12 *P/ren*. Nucleotides 2127 through 2269 of DLP12 were identical to the first 143 bases of the IS*3* element (49). This IS*3* sequence is part of IS*3* $\alpha_4\beta_4$. Both *argU* and IS*3* $\alpha_4\beta_4$ have been mapped near min 12.5 (14, 24, 27), and a comparison of the restriction map of that region (22) localized the left side of IS*3* $\alpha_4\beta_4$ to a 4.5-kb *Hin*dIII fragment at 579 kb on the chromosome (31). Comparison of the restriction enzyme sites on pDM1 chromosomal DNA located *argU*, DLP12, and the 143 bp of IS*3* to the same 4.5-kb *Hin*dIII fragment. Moreover, the 143 bp of IS*3* sequence adjacent to the DLP12 *P/ren* segment had the same orientation as the IS*3* $\alpha_4\beta_4$.

***argU* and DLP12 *int* occur in other *E. coli* strains and in other species.** To determine whether *argU* and DLP12 *int* exist in other *E. coli* strains or in other species, we probed genomic DNA samples for *argU* and DLP12 *int* sequences. To make an RNA probe specific for *argU*, pDC403 was linearized with *Bss*HII and transcribed by using T7 RNA polymerase; to make an RNA probe for DLP12 *int*, pDC403 was linearized with *Eco*RV and transcribed by using SP6 RNA polymerase (Fig. 1B). Genomic DNA isolated from several *E. coli* strains and from *Salmonella typhimurium*, *Enterobacter aerogenes*, *Erwinia carotavora*, *Klebsiella pneumoniae*, *Shigella boydii*, and *Vibrio harveyi* was digested with *Hin*dIII, fractionated by size on agarose, transferred to GeneScreen, and probed for *argU* or DLP12 *int* sequences.

The *argU* probe hybridized to a single restriction fragment from all the *E. coli* strains tested (B, C, K-12 C600, and F) and from *Salmonella typhimurium* and *Shigella boydii* (8.4-, 10.2-, 4.5-, 23-, 2.2-, and 9.4-kb *Hin*dIII fragments, respectively) (Fig. 5). Identical results were obtained when the Southern blots were probed for DLP12 *int*, indicating that DLP12 *int* and *argU* are probably located on the same restriction fragment. (One additional fragment of about 22 kb from *E. coli* C hybridized to the DLP12 *int* probe.) These results are consistent with the hypothesis that the *argU*/DLP12 *int* gene arrangement found in *E. coli* K-12 is also

FIG. 3. Comparison of DLP12 sequences with phage sequences. (A) The deduced DLP12 Int protein sequence (top line) compared with P22 integrase (34). Identical amino acids are marked by dashes, conservative changes (13) are shown by a colon, and where nonconservative differences occur, the residue found in P22 Int is indicated. (B) Comparison of the presumptive DLP12 *att* half-site with P22 *attP* sequences (33). The solid overline designates the nucleotides that encode the *argU* tRNA, the dashed overline designates the 3′ end of the *int* ORF, and the solid underline indicates the P22 *att* core. Potential core Int recognition sequences are designated by dashed arrow, and the sequences resembling the IHF consensus sequence are indicated by a solid arrow. Bases identical in *argU*/DLP12 and P22 *attP* are repeated between the sequences. (C) Comparison of DLP12 nucleotide sequences (top line) with phage sequences (12, 34). Identical bases are marked by dahses, and where differences occur, the substitution is indicated. (C-1) DLP12 *xis* and P22 *xis*. The DLP12 *xis* comparison runs to the stop codon for P22 *xis*. (C-2) DLP12 *exo* and λ *exo*. (C-3) DLP12 *P/ren* and λ *P/ren*.

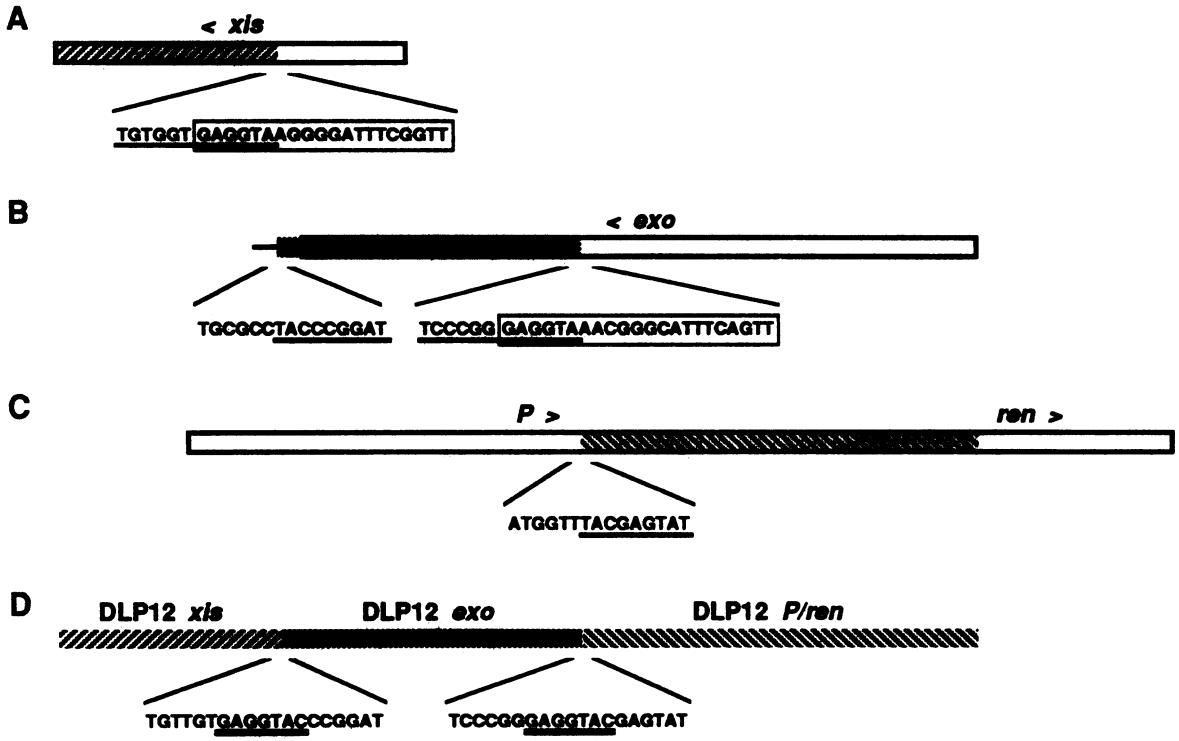


FIG. 4. Junctions of DLP12 segments and comparison with homologous segments of P22 and λ. Arrows designate transcription direction or orientation of a gene segment. In A, B, and C, phage sequences in regions homologous to DLP12 junctions are shown. Bold underlines designate the sequences homologous to DLP12 and therefore indicate the boundaries of the homologies. Nucleotides within boxes are related sequences found in P22 *xis* and in λ *exo*, where homology with DLP12 sequences ended. (A) P22 *xis* ORF, with the segment homologous to DLP12 *xis* crosshatched. (B) A segment of λ including the *exo* ORF (box) and adjacent nucleotides. The crosshatch indicates homology with the DLP12 *exo* region. (C) The λ *P* and *ren* ORFs. The crosshatch indicates homology with DLP12. (D) DLP12. Nucleotide sequences at the junctions are presented below the DLP12 bar. The 7-bp direct repeats are indicated by double underline.

present in other *E. coli* strains and some closely related species. We detected no *argU* or DLP12 *int* sequences in the other species tested (data not shown).

DLP12 *int* is transcribed in vivo. Whole-cell RNA was isolated from strain JM103 containing plasmid pDM1 or pBR322, fractionated by electrophoresis, transferred to GeneScreen, and probed for DLP12 *int* transcripts. Radiolabeled RNA probes were prepared by transcribing pDC403 (Fig. 1B) linearized with *Eco*RI by using T7 RNA polymerase. The principal DLP12 *int* transcripts from the strain containing pDM1 were about 1,350, 1,100, 850, and 600 nucleotides long (Fig. 6, lanes b and c). In some gels, these

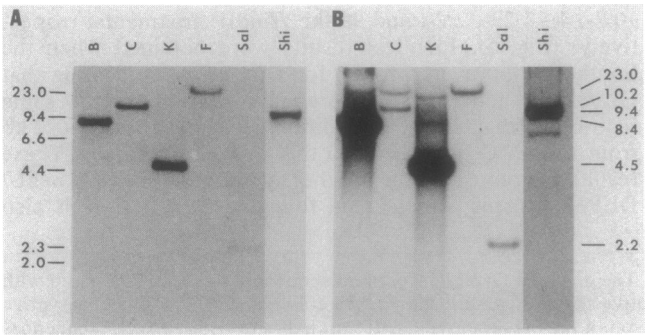


FIG. 5. Southern blots probed for *argU* or DLP12 *int* sequences. Genomic DNA from *E. coli* strains B, C, K-12, and F and from *Salmonella typhimurium* (Sal) and *Shigella boydii* (Shi) probed for *argU* (A) or DLP12 *int* (B). The sizes (in kilobases) of the restriction fragments are shown to the right. Molecular size markers (left, in kilobases) were from a *Hin*dIII digest of λ.

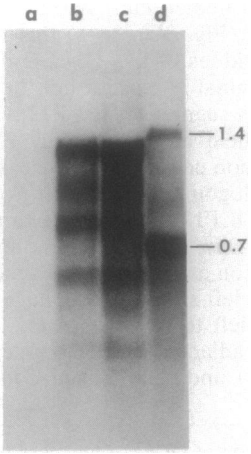


FIG. 6. Transcripts from DLP12 *int*. Northern blots of total cellular RNA extracted from strain JM103 containing pDM1 (DLP12 *int*$^+$) or strain JM103(pBR322) strains probed for DLP12 *int* sequences. Lane a, 20 μg of JM103(pBR322) RNA; lane b, 20 μg of JM103(pDM1) RNA; lane c, 40 μg of JM103(pDM1) RNA; lane d, T7 transcripts of the Riboprobe Gemini Positive control template (Promega Biotec). The sizes (in kilobases) of the T7 transcripts are shown to the right.

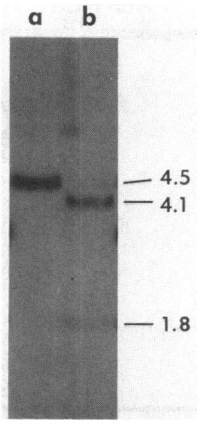


FIG. 7. Southern analysis of the DLP12 *int*::KanR insertion mutant DCK11. Genomic DNA isolated from the *int*$^+$ strain JC7623 (lane a) and the insertion strain DCK11 (lane b) was probed for DLP12 *int* sequences. Probes were transcribed from pDC403 linearized with *Bam*HI by SP6 polymerase (Fig. 1B). The sizes (in kilobases) of the restriction fragments are shown to the right.

bands appeared as doublets. Also observed was a 2.5-kb DLP12 *int* transcript, although it was barely visible. If the chromosomal DLP12 *int* is transcribed (Fig. 6, lane a), its level is much lower than in the pDM1-containing strain.

Construction and characterization of a DLP12 *int*::KanR mutant. To determine whether DLP12 *int* is essential for cell growth, site-directed insertion mutagenesis (29, 52) was performed by introducing a gene for kanamycin resistance (KanR) into the open reading frame of DLP12 *int* on the chromosome. The 1.4-kb *Pst*I KanR fragment from pUC4K (50) was subcloned into pDC1 cut with *Nsi*I, and the resulting plasmid was linearized by digestion with *Bam*HI and transformed into the DLP12 *int*$^+$ *recBC sbcB* strain JC7623. Transformants were selected for Kan^r and screened for ampicillin resistance. Of five Kan^r ampicillin-sensitive colonies chosen for Southern analysis, one had a chromosomal insertion of KanR into DLP12 *int* (Fig. 7). The 1.4-kb Kan^r cassette contains one *Hin*dIII site, and its insertion into DLP12 *int*, contained on a 4.5-kb *Hin*dIII chromosomal fragment, resulted in the disappearance of the wild-type 4.5-kb *Hin*dIII fragment and the appearance of 4.1- and 1.8-kb DLP12 *int*-containing *Hin*dIII fragments, which is in agreement with the predicted sizes. The DLP12 *int*::KanR strain was designated DCK11.

The DLP12 *int*::KanR strain (DCK11) grew at the same rate as the parent strain in rich medium at 18, 37, and 43°C. At 43°C, both strains grew poorly. Both strains were equally sensitive to UV, and both grew on minimal medium supplemented with the requirements for the parent strain JC7623.

DISCUSSION

The *argU* gene of *E. coli* is located, in clockwise orientation, at 577.5 kb on the *E. coli* chromosome restriction map. Cryptic prophage sequences spanning the 2 kb immediately downstream of *argU* consist of segments similar to the phage P22 *int* gene, a portion of the P22 *xis* gene, and portions of the *exo*, *P*, and *ren* genes of λ. This cryptic prophage was designated DLP12. Similarity of a portion of the DLP12 Int deduced amino acid sequence to the family of Int recombinases has been observed independently (53). These authors predicted that this region may be the end of a cryptic prophage.

It is probable that other *E. coli* K-12 sublines contain similar DLP12 regions. The DNA sequence of the *argU*/DLP12 region (Fig. 2) was determined from a 4.5-kb *Hin*dIII fragment of chromosomal DNA subcloned from plasmid pLC22-8, originally prepared from a strain of the 58 subline (3, 10). Probes for DLP12 *int* hybridized to a 4.5-kb *Hin*dIII chromosomal fragment of K-12 strain C600, a derivative of subline 167 (3). The *argU*/DLP12 region can be localized to a *Hin*dIII fragment of about 4.5 kb at 575 to 579.5 kb by inspection of the physical map (31) prepared from strain W3110, a subline of W1485 (3). Thus, the genomes of three commonly used *E. coli* K-12 strains, each derived from a different early isolate, all contain a 4.5-kb *Hin*dIII genomic DNA fragment with *argU*/DLP12 *int*.

In addition, the genomes of several different *E. coli* strains and of *Salmonella typhimurium* and *Shigella boydii* contain sequences highly similar to *argU* and DLP12 *int*. For each strain or species, the region homologous to *argU* occurs on a *Hin*dIII fragment of the same size as the fragment homologous to DLP12 *int*. It is likely that each of these contains *argU* and DLP12 *int* genes in the same arrangement as the K-12 strain sequenced in this report. (The DLP12 *int* probe hybridized to two *Hin*dIII fragments of *E. coli* C genomic DNA. Although there might be two DLP12 *int*-like regions in this strain, this result could also be accounted for by the presence of a *Hin*dIII site within the *E. coli C* DLP12 *int*.) Thus, the *argU*/DLP12 *int* gene organization originated in nature possibly prior to the divergence of *E. coli*, *Salmonella typhimurium*, and *Shigella boydii*. This conclusion is further strengthened by the fact that the *Shigella boydii* strain used was a recent clinical isolate. Perhaps the *argU*/DLP12 *int* sequence has been retained because it performs a useful function.

The 25-kb region clockwise of min 12 on the *E. coli* chromosome contains several homologies to lambdoid phages. Figure 8, which extends the map of Redfield and Campbell (42), demonstrates that the DLP12 half-*att* site, *int*, *xis*, *exo*, and *P*/*ren* homology is located at 577 to 579.5 kb. The *qsr'* cryptic prophage consists of genes analogous to λ *Q*, *S*, and *R*, a *cos* site, and at least part of Nu1 and contains IS*5*-*nmpC* (4, 9, 25). Comparison of the IS*5*-*nmpC* restriction map (4, 25) with the physical map of the chromosome (31) places this region at 587 to 590 kb. Further clockwise on the chromosome are the *appY* (or M5), *ompT*, and *envY* genes (2, 19, 35, 44). Analysis of the *appY* sequence (2, 30) reveals a 151-bp sequence 5' of the gene which is 84% identical to a portion of the λ *orf-194* (12). The regulatory region of *ompT* (20, 21) contains 38 bp which are identical to 38 bp of *argU*, including the half-*att* core region. The *appY*-*ompT*-*envY* region (2, 21, 30, 36) is located at 596 to 600 kb (Fig. 8). Thus, this 25-kb region contains a half-*att* site on the left, an *int xis exo* and *P*/*ren* region, *qsr'*, a region homologous to λ *orf-194*, and a potential *att* site on the right (Fig. 8), all in the same order and orientation as the lambdoid prophage counterparts (8, 12). It is conceivable that this entire region evolved from an ancestral prophage, and we propose that all the phage homologs be designated DLP12. Inasmuch as *Salmonella typhimurium* and *Shigella boydii* contain a homolog of DLP12 *int* and presumably a similar *argU* (host)-*int* (phage) gene arrangement, we imagine that this primordial prophage was present in ancestral members of the family *Enterobacteriaceae*. Alternatively, the same distribution of the *argU*-DLP12 *int* region could have resulted from independent insertions of phages at the same site, provided that phages with that insertion specificity were common enough.

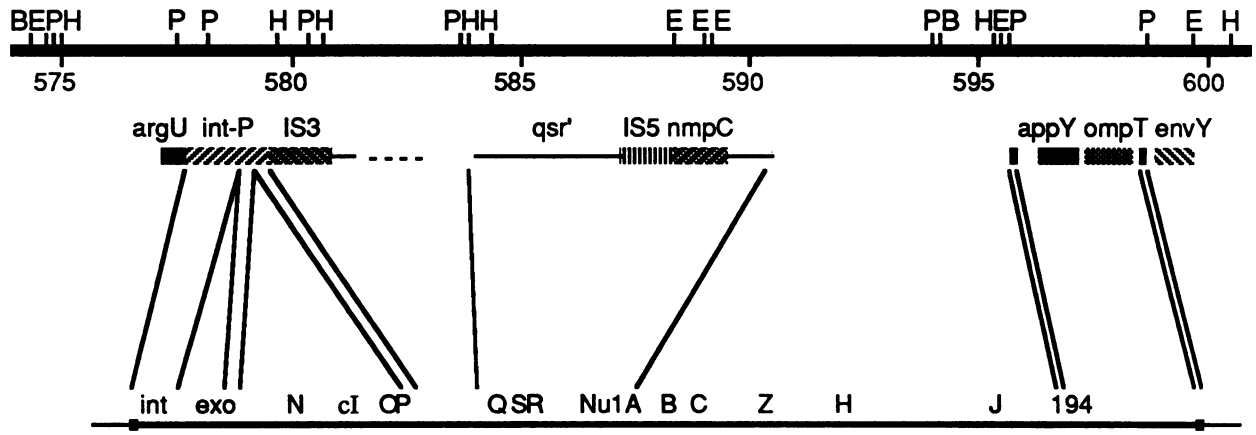


FIG. 8. Physical map of the DLP12 cryptic prophage and comparison to λ prophage. The top line represents *E. coli* chromosomal DNA, with physical map coordinates given in kilobases (31) below the line. Restriction enzymes: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I. The locations of genes and cryptic prophage segments are identified as boxes below the line. The box labeled int-P includes DLP12 *int*, *xis*, *exo*, and *P/ren*. The bottom line is a λ prophage genetic map (8, 12). The lines running between the *E. coli* and λ maps identify the regions of sequence homology. The exact extent of homology in the *qsr'* region is unknown.

It is very interesting that several *E. coli* phages and plasmids integrate into known and presumptive tRNA genes (43). Examples include P4 integration into the *E. coli* C *leuX* gene ($tRNA^{Leu}_{CAA}$) (41, 48, 54) and P22 into a *Salmonella typhimurium* sequence which is probably a threonine tRNA gene ($tRNA^{Thr}_{CGU}$) (41). The *thrW* gene of *E. coli* ($tRNA^{Thr}_{CGU}$) (11, 16) is identical to the $tRNA^{Thr}_{CGU}$ gene of *Salmonella typhimurium*, and inasmuch as the *E. coli* attachment site for P22 has been mapped near min 6 (26), the location of *thrW* (11, 16), we propose that P22 also integrates into this tRNA gene in *E. coli*. At least two additional tRNA gene sequences ($tRNA^{Thr}_{CGU}$ of *Pseudomonas aeruginosa* [11, 40] and $tRNA^{Thr}_{UCU}$ of *Mycoplasma mycoides* [45]) share extensive homology with the P22 *attB* core, suggesting that they could also serve as attachment sites. In P22, P4, and DLP12, the attachment core ends precisely at the position of the gene which encodes the 3′ end of the tRNA. Integration leaves the mature tRNA coding sequences intact and replaces the 3′ end of the tRNA gene with phage integrase genes. Pierson and Kahn point out (41) that attachment site locations in tRNA genes could provide advantages to the host or phage.

DLP12 may have contributed to the evolution of lambdoid phages. The modular theory of bacteriophage evolution (6, 9, 47) suggests that a bacterial host carrying prophage or defective prophage sequences could serve as a reservoir for the exchangeable modules. DLP12 is especially interesting because it contains both P22 and λ "modules," even though they are highly diverged and grow on different hosts.

It is possible that the present DLP12 structure has undergone many changes. These might include recombinational reshuffling with other phage genomes and homologous recombination which interspersed host genes within the prophage. Alternatively, genes such as *ompT* and *appY* might have been homologous to prophage genes in the corresponding positions of the prophage. Insertion elements cause deletions and rearrangements in nearby regions (15). The directly repeated sequences found at the DLP12 junctions and in the analogous regions of the phage genomes may have played a role in genetic rearrangements.

ACKNOWLEDGMENTS

We thank J. W. Zyskind and I. J. Molineux for strains and S. M. Payne for a DNA preparation.

This work was supported by Public Health Service grant GM34471 from the National Institutes of Health and, in part, by American Cancer Society grant MV-429M, Welch Foundation grant F949, and University of Texas Project Quest.

LITERATURE CITED

1. **Argos, P., A. Landy, K. Abremski, J. B. Egan, E. Haggard-Ljungquist, R. H. Hoess, M. L. Kahn, B. Kalionis, S. V. L. Narayana, L. S. Pierson III, N. Sternberg, and J. M. Leong.** 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. EMBO J. **5**:433–440.
2. **Atlung, T., A. Nielsen, and F. G. Hansen.** 1989. Isolation, characterization, and nucleotide sequence of *appY*, a regulatory gene for growth-phase-dependent gene expression in *Escherichia coli*. J. Bacteriol. **171**:1683–1691.
3. **Bachmann, B. J.** 1972. Pedigrees of some mutant strains of *Escherichia coli* K-12. Bacteriol. Rev. **36**:525–557.
4. **Blasband, A. J., W. R. Marcotte, Jr., and C. A. Schnaitman.** 1986. Structure of the *lc* and *nmpC* outer membrane porin protein genes of lambdoid bacteriophage. J. Biol. Chem. **261**: 12723–12732.
5. **Blum, P., D. Holzschu, H.-S. Kwan, D. Riggs, and S. Artz.** 1989. Gene replacement and retrieval with recombinant M13mp bacteriophages. J. Bacteriol. **171**:538–546.
6. **Botstein, D.** 1980. A theory of modular evolution for bacteriophages, p. 484–491. *In* P. Palese and B. Roizman (ed.), Genetic variation of viruses. New York Academy of Sciences, New York.
7. **Bovre, K., H. A. Lozeron, and W. Szybalski.** 1971. Techniques of RNA-DNA hybridization in solution for the study of viral transcription. Methods Virol. **5**:271–292.
8. **Campbell, A.** 1971. Genetic structure, p. 13–44. *In* A. D. Hershey (ed.), The bacteriophage lambda. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
9. **Campbell, A.** 1988. Phage evolution and speciation, p. 1–14. *In* R. Calendar (ed.), The bacteriophages, vol. 1. Plenum Publishing Corp., New York.
10. **Clarke, L., and J. Carbon.** 1976. A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire *E. coli* genome. Cell **9**:91–99.
11. **Dalrymple, B., and J. S. Mattick.** 1986. Genes encoding threonine tRNAs with the anticodon CGU from *Escherichia coli* and *Pseudomonas aeruginosa*. Biochem. Int. **13**:547–553.
12. **Daniels, D. L., J. L. Schroeder, W. Szybalski, F. Sanger, A. R. Coulson, G. F. Hong, D. F. Hill, G. B. Petersen, and F. R. Blattner.** 1983. Appendix II: complete annotated lambda sequence, p. 519–676. *In* R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg (ed.), Lambda II. Cold Spring Harbor

Laboratory, Cold Spring Harbor, N.Y.
13. **Dayhoff, M. O., R. M. Schwartz, and B. L. Orcott.** 1978. A model of evolutionary change in proteins, p. 345–352. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5, Suppl. 3. National Biomedical Research Foundation, Washington, D.C.
14. **Deonier, R. C., G. R. Oh, and M. Hu.** 1977. Further mapping of IS*2* and IS*3* in the *lac-purE* region of the *Escherichia coli* K-12 genome: structure of the F-prime ORF203. J. Bacteriol. **129:** 1129–1140.
15. **Deonier, R. C., K. Yun, and M. Kuppermann.** 1983. γδ-Mediated deletions of chromosomal segments on F-prime plasmids. Mol. Gen. Genet. **190:**42–50.
16. **Deutch, A. H., K. E. Rushlow, and C. J. Smith.** 1984. Analysis of the *Escherichia coli proBA* locus by DNA and protein sequencing. Nucleic Acids Res. **12:**6337–6355.
17. **Fournier, M. J., and H. Ozeki.** 1985. Structure and organization of the transfer ribonucleic acid genes of *Escherichia coli* K-12. Microbiol. Rev. **49:**379–397.
18. **Garcia, G. M., P. K. Mar, D. A. Mullin, J. R. Walker, and N. E. Prather.** 1986. The *E. coli dnaY* gene encodes an arginine transfer RNA. Cell **45:**453–459.
19. **Gayda, R. C., H. Avni, P. E. Berg, and A. Markovitz.** 1979. Outer membrane protein *a* and other polypeptides regulate capsular polysaccharide synthesis in *E. coli* K-12. Mol. Gen. Genet. **175:**325–332.
20. **Gordon, G., R. C. Gayda, and A. Markovitz.** 1984. Sequence of the regulatory region of *ompT*, the gene specifying major outer membrane protein *a* (3b) of *Escherichia coli* K-12: implications for regulation and processing. Mol. Gen. Genet. **193:**414–421.
21. **Grodberg, J., M. D. Lundrigan, D. L. Toledo, W. F. Mangel, and J. J. Dunn.** 1988. Complete nucleotide sequence and deduced amino acid sequence of the *ompT* gene of *Escherichia coli* K-12. Nucleic Acids Res. **16:**1209.
22. **Hadley, R. G., M. Hu, M. Timmons, K. Yun, and R. C. Deonier.** 1983. A partial restriction map of the *proA-purE* region of the *Escherichia coli* K-12 chromosome. Gene **22:**281–287.
23. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11:**2237–2255.
24. **Henson, J. M., H. Chu, C. A. Irwin, and J. R. Walker.** 1979. Isolation and characterization of *dnaX* and *dnaY* temperature-sensitive mutants of *Escherichia coli*. Genetics **92:**1041–1059.
25. **Highton, P. J., Y. Chang, W. R. Marcotte, Jr., and C. A. Schnaitman.** 1985. Evidence that the outer membrane protein gene *ompC* of *Escherichia coli* K-12 lies within the defective *qsr'* prophage. J. Bacteriol. **162:**256–262.
26. **Hoppe, I., and J. Roth.** 1975. Specialized transducing phages derived from *Salmonella* phage P22. Genetics **76:**633–654.
27. **Hu, S., E. Ohtsubo, and N. Davidson.** 1975. Electron microscope heteroduplex studies of sequence relations among plasmids of *Escherichia coli*: structure of F13 and related F-primes. J. Bacteriol. **122:**749–763.
28. **Ikemura, T.** 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. J. Mol. Biol. **146:**1–21.
29. **Jasin, M., and P. Schimmel.** 1984. Deletion of an essential gene in *Escherichia coli* by site-specific recombination with linear DNA fragments. J. Bacteriol. **159:**783–786.
30. **Kemp, E. H., N. P. Minton, and N. H. Mann.** 1987. Complete nucleotide sequence and deduced amino acid sequence of the M5 polypeptide gene of *Escherichia coli*. Nucleic Acids Res. **15:**3924.
31. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50:**495–508.
32. **Kushner, S. R., H. Nagaishi, A. Templin, and A. J. Clark.** 1971. Genetic recombination in *Escherichia coli*: the role of exonuclease I. Proc. Natl. Acad. Sci. USA **68:**824–827.
33. **Leong, J. M., S. Nunes-Düby, C. F. Lesser, P. Youderian, M. M. Susskind, and A. Landy.** 1985. The φ80 and P22 attachment sites: primary structure and interaction with *Escherichia coli* integration host factor. J. Biol. Chem. **260:**4468–4477.
34. **Leong, J. M., S. E. Nunes-Düby, A. B. Oser, C. F. Lesser, P. Youderian, M. M. Susskind, and A. Landy.** 1986. Structural and regulatory divergence among site-specific recombination genes of lambdoid phage. J. Mol. Biol. **189:**603–616.
35. **Lundrigan, M. D., and C. F. Earhart.** 1984. Gene *envY* of *Escherichia coli* K-12 affects thermoregulation of major porin expression. J. Bacteriol. **157:**262–268.
36. **Lundrigan, M. D., M. J. Friedrich, and R. J. Kadner.** 1989. Nucleotide sequence of the *Escherichia coli* porin thermoregulatory gene *envY*. Nucleic Acids Res. **17:**800.
37. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
38. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12:**7035–7056.
39. **Mullin, D. A., G. M. Garcia, and J. R. Walker.** 1984. An *E. coli* DNA fragment 118 base pairs in length provides *dnaY*$^+$ complementing activity. Cell **37:**669–674.
40. **Pasloske, B. L., B. B. Finlay, and W. Paranchych.** 1985. Cloning and sequencing of the *Pseudomonas aeruginosa* PAK pilin gene. FEBS Lett. **183:**408–412.
41. **Pierson, L. S., III, and M. L. Kahn.** 1987. Integration of satellite bacteriophage P4 in *Escherichia coli*. DNA sequences of the phage and host regions involved in site-specific recombination. J. Mol. Biol. **196:**487–496.
42. **Redfield, R. J., and A. M. Campbell.** 1987. Structure of cryptic λ prophages. J. Mol. Biol. **198:**393–404.
43. **Reiter, W.-D., P. Palm, and S. Yeats.** 1989. Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. Nucleic Acids Res. **17:**1907–1914.
44. **Rupprecht, K. R., G. Gordon, M. Lundrigan, R. C. Gayda, A. Markovitz, and C. Earhart.** 1983. *ompT*: *Escherichia coli* K-12 structural gene for protein a (3b). J. Bacteriol. **153:**1104–1106.
45. **Samuelson, J., Y. S. Grindy, F. Lustig, T. Boren, and U. Lagerkvist.** 1987. Apparent lack of discrimination in the reading of certain codons in *Mycoplasma mycoides*. Proc. Natl. Acad. Sci. USA **84:**3166–3170.
46. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
47. **Susskind, M. M., and D. Botstein.** 1978. Molecular genetics of bacteriophage P22. Microbiol. Rev. **42:**385–413.
48. **Thorbjarnardóttir, S., T. Dingermann, T. Rafnar, O. S. Andrésson, D. Söll, and G. Eggertsson.** 1985. Leucine tRNA family of *Escherichia coli*: nucleotide sequence of the *supP*(Am) suppressor gene. J. Bacteriol. **161:**219–222.
49. **Timmerman, K. P., and C.-P. D. Tu.** 1985. Complete sequence of IS*3*. Nucleic Acids Res. **13:**2127–2139.
50. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19:**259–268.
51. **Walker, J. R.** 1987. A transfer RNA implicated in DNA replication, p. 261–268. *In* M. Inouye and B. S. Dudock (ed.), Molecular biology of RNA: new perspectives. Academic Press, Inc., San Diego.
52. **Winans, S. C., S. J. Elledge, J. H. Krueger, and G. C. Walker.** 1985. Site-directed insertion and deletion mutagenesis with cloned fragments in *Escherichia coli*. J. Bacteriol. **161:**1219–1221.
53. **Yagil, E., S. Dolev, J. Obertor, N. Kislev, N. Ramaiah, and R. A. Weisberg.** 1989. Determinants of site-specific recombination in the lambdoid coliphage HK022: an evolutionary change in specificity. J. Mol. Biol. **207:**695–717.
54. **Yoshimura, M., H. Inokuchi, and H. Ozeki.** 1984. Identification of transfer RNA suppressors in *Escherichia coli*. IV. Amber suppressor Su^{+6}, a double mutant of a new species of leucine tRNA. J. Mol. Biol. **177:**627–644.