Conservation and DNA Sequence Arrangement of the DNA Polymerase I Gene Region from *Klebsiella aerogenes*, *Klebsiella pneumoniae* and *Escherichia coli*

Kevin Struhl† and Ronald W. Davis

Department of Biochemistry
Stanford University School of Medicine
Stanford, Calif. 94305, U.S.A.

Collections of viable molecular hybrids of bacteriophage λ DNA and of *Klebsiella aerogenes* or *Klebsiella pneumoniae* DNA were constructed *in vitro* by the covalent joining of DNA fragments *via* the cohesive ends generated by *Eco*RI endonuclease cleavage. From each of these two pools of hybrids, a phage containing the gene coding for DNA polymerase I (*polA*) was selected. Functional *Klebsiella* DNA polymerase I is produced when either hybrid phage is grown in *Escherichia coli* hosts. Deletion mutants of bacteriophage λ hybrids containing the *K. pneumoniae* polymerase I gene generated during lytic growth in *E. coli* have been isolated. One particular mutation (a $3\cdot2 \times 10^3$-base deletion entirely within the cloned *K. pneumoniae* DNA fragment) was isolated at an extremely high frequency $(0\cdot5 \times 10^{-4})$. The location of the *polA* gene on the cloned DNA fragment has been determined by correlating physical structures and genetic expression of the deletion mutants.

Heteroduplex analysis indicates extensive homology but not complete co-linearity in the DNA sequences of the polymerase genes and adjacent regions from *K. aerogenes*, *K. pneumoniae* and *E. coli*. Using very stringent hybridization conditions, it has been determined that most if not all of the *polA* gene is highly conserved among these organisms; the adjacent sequences are significantly less conserved. DNA sequences homologous to the cloned *polA*-containing fragments from *K. aerogenes* and *K. pneumoniae* were not detected in the DNA of *Bacillus subtilis* or *Myxococcus xanthus*. Surprisingly, at two or three locations, a given DNA strand of the *K. aerogenes* fragment is homologous to both DNA strands of the *K. pneumoniae* fragment. Some of these locations map close to breakpoints between highly conserved and less conserved regions. Models incorporating these observations are considered.

1. Introduction

The enteric bacteria are closely related. In general, genes from one of these organisms are functionally expressed in another (Baron *et al.*, 1968). The genomes are quite homologous (Brenner *et al.*, 1969). The genetic maps are similar though differences are known (Middleton, 1971; Streicher *et al.*, 1975; Bachmann *et al.*, 1976). However, recombination between genes of different enteric bacteria is infrequent though detectable (Middleton, 1971; Tyler & Goldberg, 1976); this suggests that equivalent

† Present address: Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England.

16

local regions of the genomes may not be as conserved in DNA sequence as the broad overall organization. These observations are consistent with the supposition that these enteric organisms have evolved divergently from a common primordial ancestor.

To examine the nature of conservation at the molecular level, we have compared the DNA sequences which contain the DNA polymerase I genes and the adjacent regions from *Klebsiella aerogenes*, *Klebsiella pneumoniae*, and *Escherichia coli*. DNA polymerase I is enzymologically and antigenically conserved across many bacterial genera (Tafler *et al.*, 1973). In this paper, we describe the isolation of the genes for DNA polymerase I (*polA*) from pools of bacteriophage λ hybrids containing DNA fragments of *K. aerogenes* or *K. pneumoniae* generated by *Eco*RI endonuclease cleavage. The selection for λ*polA* phages depends upon the inability of *red*⁻ derivatives of λ to grow lytically on DNA polymerase I-deficient, *E. coli* hosts (Gottesman *et al.*, 1973). The cloning vector, λgt1-λB, and all λgt1 hybrid phage are deleted for a portion of the *red* genes. However, a hybrid phage which expresses functional *Klebsiella* DNA polymerase I grows lytically on an *E. coli* host deficient in DNA polymerase I (Struhl & Davis, 1976). In this manner Kelley *et al.* (1977) isolated the gene for DNA polymerase I from *E. coli*. Using the hybrid phages containing the cloned DNA polymerase I genes from *K. aerogenes*, *K. pneumoniae* and *E. coli* we have examined the conservation of the DNA sequences. Conservation has been studied directly by heteroduplex analysis of the cloned *polA* genes, and indirectly by nucleic acid hybridization of the cloned DNAs to total DNAs of organisms from which the polymerase I genes have not been cloned.

2. Materials and Methods

(a) Reagents

Activated calf thymus DNA, purified DNA polymerase I from *E. coli*, and antibody prepared against it were gifts from Dennis Uyemura and Janice Chien. Other reagents have been described (Thomas *et al.*, 1974; Thomas & Davis, 1975).

(b) Bacteria, phage and DNA

E. coli and λ strains used in this paper have been described (Thomas *et al.*, 1974; Panasenko *et al.*, 1977). The strain of *K. aerogenes* was MK9000 (*hutC*515 P1ˢ) from the collection of Boris Magasanik (Streicher *et al.*, 1975). The strain of *K. pneumoniae*, KP1, was obtained from Stanley Streicher (Streicher *et al.*, 1971). *E. coli*, *Saccharomyces cerevisiae*, *Dictyostelium discoideum*, *Drosophila melanogaster*, and *Bacillus subtilis* DNAs were obtained, respectively, from John Cameron, Benjamin Hall, Allan Jacobson, David Finnegan and Grace Chi. Bacterial and λ DNAs were prepared by the procedure of Cameron *et al.* (1975). The structures of the phages used in this paper are shown in Fig. 1. The original isolates are given the following culture collection numbers: λgt1-*Kp polA′* = NNλ60, λgt1-*Kp polA* = NNλ61, and λgt1-*Ka polA* = NNλ62.

(c) Preparation of λgt hybrid pools containing DNA fragments from K. aerogenes and from K. pneumoniae

Hybrid DNA molecules were made by separately cleaving λgt1-λB vector DNA and bacterial DNA with *Eco*RI restriction endonuclease, mixing, and covalently joining the fragments with *E. coli* DNA ligase. Hybrid DNA molecules were tranfected into CaCl₂-treated cells of strain SF8 (C600 *hsr*⁻ *hsm*⁻ *recBC*⁻ *lop*11) according to Mandel & Higa (1970) as modified by Cameron *et al.* (1975). Complete cleavage of λgt1-λB DNA with
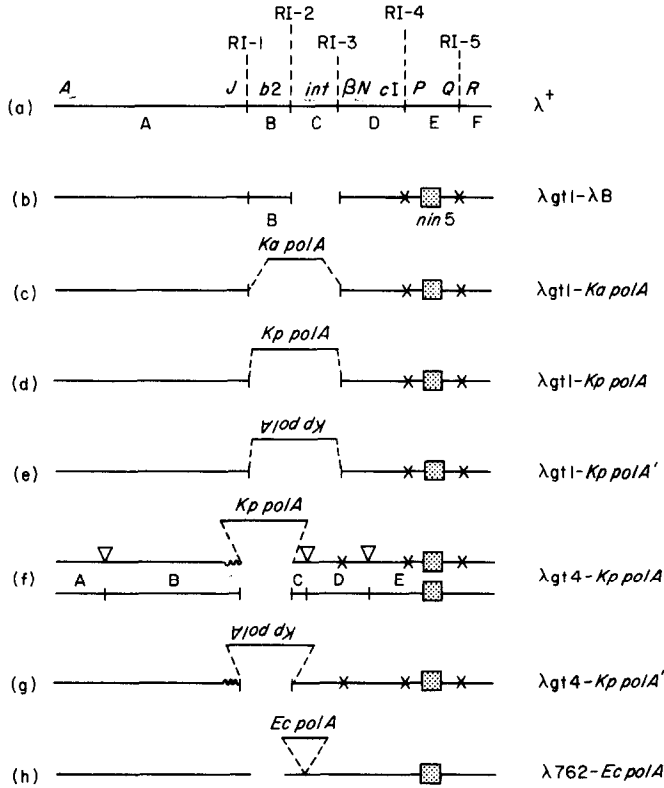
FIG. 1. Structures of hybrid phages. (a) Structure of wild-type λ. λ genes are listed above the horizontal lines. *Eco*RI fragments are listed below the horizontal lines; the five *Eco*RI recognition sites are designated as vertical lines. (b) Structure of the vector λgt1-λB. The *Eco*RI-λC fragment is deleted and the *Eco*RI-4 and *Eco*RI-5 sites are mutated. (c), (d) and (e) Structures of λgt1 hybrids containing the *K. aerogenes* or the *K. pneumoniae polA Eco*RI-generated DNA fragment. The orientation of the cloned fragment with respect to the vector (indicated by *polA* or by *Alop*) is represented arbitrarily. (f) and (g) Structures of λgt4 hybrids containing the *K. pneumoniae polA* fragment. The orientation of the cloned fragment is shown with respect to the λgt1 hybrids. The λgt4 vector lacks the *Eco*RI-λB fragment and has a mutated *Eco*RI-3 site. The *Bam*HI sites of λgt4 are indicated by the open triangles in (f). The restriction spectra of λgt4 DNA produced by cleavage with *Bam*HI and *Eco*RI are indicated between (f) and (g). (h) Structure of the hybrid containing the cloned *E. coli polA* fragment as determined by Kelley *et al.* (1977). The λ762 vector lacks the *Eco*RI-λB fragment. The *E. coli* fragment was cloned as a DNA fragment generated by *Hind*III endonuclease; it was inserted at the *Hind*III site at 0·570 on the λ map (just to the left of *att*). All hybrid phages contained the *cI857* and the *nin5* mutations.

*Eco*RI endonuclease produces three fragments: the left and right end fragments of λgt1 and the B fragment of λ DNA (Thomas & Davis, 1975). Though the left and the right ends of λgt1 DNA contain all the essential genetic information for lytic growth of λ, a molecule consisting solely of these λgt1 ends is too short to produce a plaque-forming phage. Viable molecular hybrids, therefore, must contain an insertion of DNA between the λgt1 end fragments. Insertion of foreign DNA is favored over re-insertion of the B fragment of λ DNA when appropriate concentrations of vector and foreign DNA are used. Upon transfection of hybrid DNA preparations, 50 to 95% of the plaques result from viable molecular hybrids. Phage from such transfections are combined to form a hybrid pool and

subsequently K-modified by passage on C600 hsr^- hsm^+. Such a K-modified phage pool can then be used to infect essentially any strain of $E.$ $coli$ K12. These methods are described by Thomas et $al.$ (1974).

(d) Enzyme reactions

DNA polymerase I assays were performed as described by Lehman & Chien (1973). All restriction endonuclease mixtures contained 100 μg gelatin/ml and 6 mM-β-mercaptoethanol. EcoRI and BamH1 digests were performed in 50 mM-Tris (pH 7·4), 100 mM-NaCl, 10 mM-MgCl$_2$. $Hinc$II and AvaII reactions were carried out in 10 mM-Tris (pH 7·4), 50 mM-NaCl, 7 mM-MgCl$_2$. Cleavage by HpaI and AvaI endonucleases was done in 10 mM-Tris (pH 7·4), 10 mM-NaCl, 10 mM-MgCl$_2$.

(e) Agarose gel electrophoresis

DNAs were cleaved with restriction endonuclease, heated to 70°C for 2 min, and made 10% in sucrose. Before 1976, samples were subjected to gel electrophoresis in tubes containing 0·7% (w/v) agarose according to Thomas & Davis (1975). At later dates, DNA fragments were separated in slabs of agarose as described by Struhl & Davis (1980). Fragment lengths were calibrated using restriction fragments of λ DNA as standards. The length of λ DNA used is 49·0 kb† (Philippsen et $al.$, 1978).

(f) Heteroduplex analysis

Heteroduplexes were made and then mounted for electron microscopy as described by Davis et $al.$ (1971). The analysis of heteroduplex molecules was performed in a manner similar to that of Ferguson & Davis (1975). Measurements of the double-stranded lengths in the insert region of a given molecule were normalized to the measured lengths of the EcoRI left and right duplex end fragments of λgt1 in that molecule. The lengths of these λgt1 end fragments were taken to be 21·8 kb for the left end and 14·0 kb for the right end (P. Philippsen & R. Davis, unpublished results). The total length of the single-stranded regions in a given heteroduplex was calculated using the known total length of the insert region and the measured total double-stranded length of the insert region of that heteroduplex. Measurements of individual single-strand regions in a given molecule were normalized to the total length of the single-stranded regions in that molecule.

(g) Isolation and physical mapping of deletion mutants

Deletion mapping was the experimental strategy employed to localize the $K.$ $pneumoniae$ $polA$ gene on the cloned DNA fragment. The procedures closely followed those described for mapping the yeast $his3$ gene (Struhl & Davis, 1980). The Kp $polA$ fragment in λgt1-Kp $polA$ DNA was transferred to λgt4 DNA (Panasenko et $al.$, 1977) by cleaving with EcoRI endonuclease and then ligating with $E.$ $coli$ DNA ligase. Because λgt4 hybrids containing the Kp $polA$ DNA fragment have a genome size equal in length to that of wild type λ, such recombinants were identified by plaque formation on GL1, an $E.$ $coli$ strain which does not plate deletion mutants of λ (Emmons et $al.$, 1975). DNAs from 12 prospective recombinant phages were prepared by the rapid lysate method of Cameron et $al.$ (1977). The EcoRI restriction spectra revealed bands corresponding only to the λgt4 end fragments and to the Kp $polA$ fragment. Electrophoretic separation of DNA fragments following cleavage by HpaI endonuclease indicated 2 classes of recombinants; these represent the Kp $polA$ DNA fragment cloned in either of the two possible orientations with respect to the λgt4 vector. The orientation of the Kp $polA$ DNA fragment in each class of λgt4 hybrid relative to λgt1-Kp $polA$ was determined by analysis of heteroduplexes. Independently derived deletion mutants of λgt4-Kp $polA$ and λgt4-Kp $polA'$ were isolated by virtue of their resistance to chelating agents in a manner described previously (Parkinson & Huskey, 1971; Struhl & Davis, 1980). Some deletion mutants were selected for $polA$

† Abbreviation used: kb, kilobase ($\times 10^3$ base-pairs).

function; others were not. Because the deletion mutants were *red*[+], the test for *polA* function was the ability of H560 (*polA*[-]) cells lysogenized by these derivatives to grow in the presence of 40 μg methylmethane sulfonate/ml (Gross & Gross, 1969). Because the deletion mutants were likely to be deficient in lysogen formation, they were integrated into the H560 genome as double lysogens with λcI857S7. The double lysogens were propagated at 30°C; the *polA*[+] deletion mutants were recovered following temperature induction of the prophages.

Many deletion mutants isolated in this way are generated by *int* mediated recombination (Parkinson & Huskey, 1971). *int* mediated deletion mutants were identified by genetic tests (Struhl & Davis, 1980). In these cases, one of the deletion end points occurs at the λ attachment site (*att*) (Parkinson & Huskey, 1971; Hoess & Landy, 1978; Struhl & Davis, unpublished results). Deletions of interest have one end point at *att* and the other end point within the cloned *K. pneumoniae* DNA fragment; therefore they remove one of the two *Eco*RI endonuclease recognition sites of the original phage. In λgt4 DNA, a restriction endonuclease cleavage site for *Bam*H1 is located in the *int* gene, 240 base-pairs to the right of *att* (Hoess & Landy, 1978; see Fig. 1). The *Kp polA* DNA fragment is not cleaved by *Bam*H1 (data not shown). Therefore, simultaneous cleavage of these deletion mutant DNAs with *Bam*H1 and *Eco*RI endonucleases generates a novel fragment diagnostic of the end point within the cloned *Kp polA* fragment. The length of this fragment is 240 base-pairs larger than the distance from the deletion end point to the junction (*Eco*RI site) between the λgt4 left end and *Klebsiella* DNA.

Deletions localized totally within the cloned *Kp polA* fragment were mapped by heteroduplex analysis. Other deletion mutants were not mapped.

(h) Nucleic acid hybridization

A ^{32}P-labelled DNA probe was synthesized *in vitro* with DNase I and DNA polymerase I by the process of "nick translation" (Rigby *et al.*, 1977) as modified by Struhl & Davis (1980). DNA to be tested for hybridization was cleaved with *Eco*RI endonuclease and subjected to agarose gel electrophoresis. This DNA was transferred from the gel to a strip of cellulose nitrate as described by Southern (1975). A [22]P-labelled DNA probe (10[6] disints/min in 4 ml) was hybridized to the strip in heat-sealed plastic bags for 18 h at 65°C in 5 × SSC (SSC is 0·15 M-NaCl, 0·015 M-sodium citrate, pH 7·0) and 1% sodium dodecyl sulfate (Struhl & Davis, 1980). The strips were washed twice for 20 min under the hybridization conditions, twice for 20 min in 2 × SSC, dried, and autoradiographed.

3. Results

(a) Isolation of hybrids which grow on polymerase I-deficient strains

*Eco*RI fragments of the total DNA from various genomes ranging in complexity from λ to *D. melanogaster* give highly reproducible, specific banding patterns (the *Eco*RI "restriction spectrum") when electrophoretically separated in agarose gels (Fig. 2). Such restriction spectra could conceivably be used for species classification since those of organisms in the same genus, *K. aerogenes* and *K. pneumoniae*, are easily distinguishable. The number average molecular weights of the *Eco*RI fragments from these organisms correlate directly with their base compositions. A + T-rich organisms such as *S. cerevisiae* have relatively small *Eco*RI DNA fragments, while G + C-rich organisms such as *Myxococcus xanthus* have large fragments; this is probably because *Eco*RI recognizes a site composed of 67% dA and dT residues. Most of the fragments from either species of *Klebsiella* are between 1 and 14 kb in length and thus can be inserted into λgt1 to produce viable molecular hybrids.
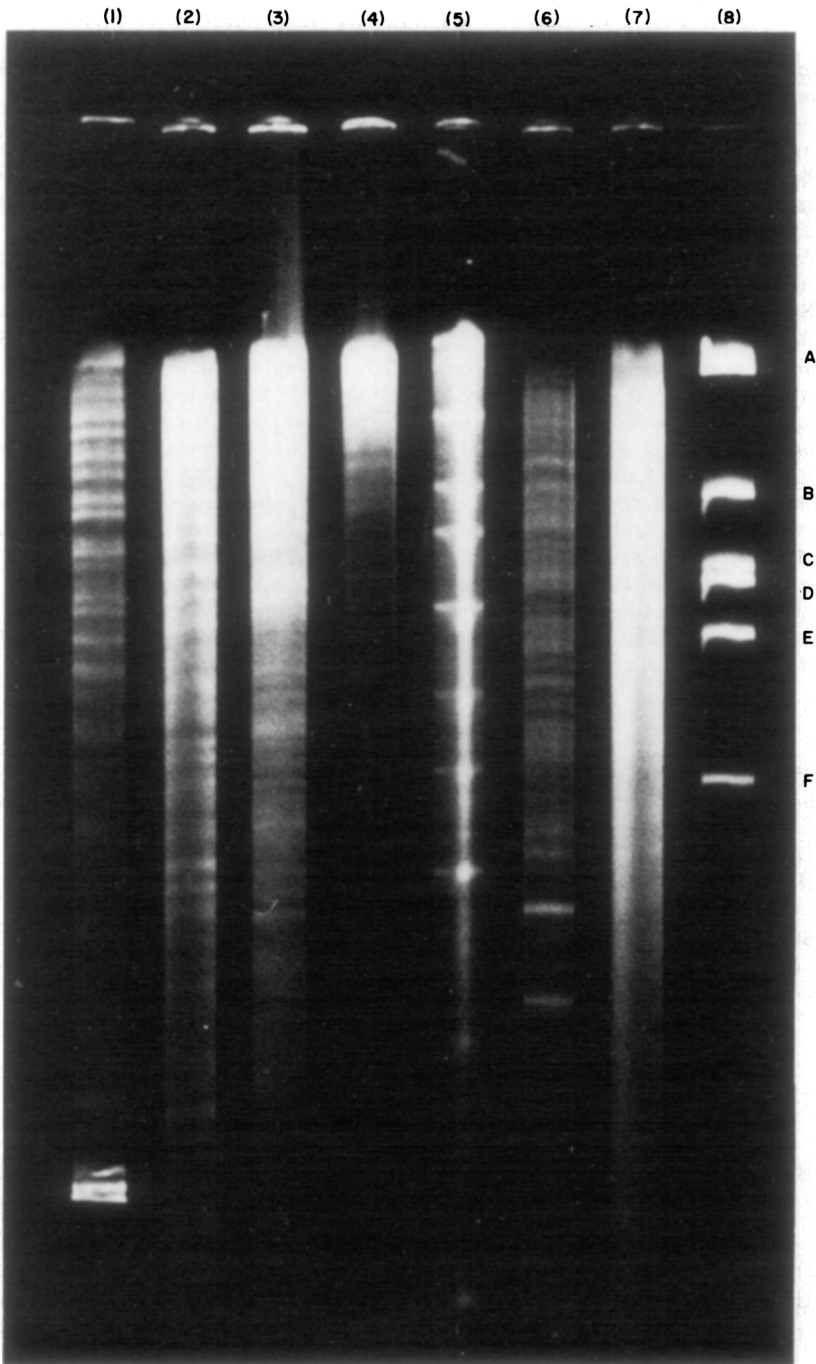
FIG. 2. *Eco*RI endonuclease restriction spectra of genomic DNAs.
Samples were subjected to electrophoresis in cylindrical tubes (0·6 cm × 14 cm) containing
0·7% agarose as described by Thomas & Davis (1975). The samples contained the following DNAs:
(1) 2 μg *E. coli* DNA; (2) 2 μg *K. aerogenes* DNA; (3) 2 μg *K. pneumoniae* DNA; (4) 2 μg *M. xanthus* DNA; (5) 3 μg *D. discoideum* DNA; (6) 3 μg *S. cerevisiae* DNA; (7) 5 μg *D. melanogaster* DNA; (8) 0·5 μg *λc*I857*S*7 DNA. *Eco*RI fragments are marked on the right-hand side of the figure.

Pools of hybrid phage were made as described in Materials and Methods using DNA from either *K. aerogenes* or *K. pneumoniae* as sources of foreign DNA. Both pools were composed of approximately 1500 independent viable hybrids. The initial vector (λgtl-λB) and all hybrid phage are *red*$^-$ and therefore are unable to grow well on *E. coli* strains lacking either DNA ligase activity (*lig*ts7) or DNA polymerase I activity (H560) (Gottesman *et al.*, 1973). However, when 10^5 K-modified phage from either hybrid pool were plated on H560 (*polA*1) at 37°C, approximately 50 large plaques were easily distinguished from a background of perhaps 1000 tiny plaques. Figure 3 shows such a plate sprayed with 100 μg ethidium bromide/ml and illuminated with a short wave, ultraviolet light source (259 nm). When a hybrid pool made from *E. coli* DNA was plated on H560, no large plaques were visible, although the background of small plaques was apparent.

(b) *Initial characterization of the hybrid phages*

Phage from five of the large plaques on H560 from the *K. pneumoniae* pool and three from the *K. aerogenes* pool were examined. Each of the eight phages plated with equal efficiency on C600 (wild-type polymerase I gene) and H560. As expected, none of the hybrids produced plaques on *lig*ts7 (Table 1). The *Eco*RI restriction spectra of the eight phage were compared. The five *K. pneumoniae* hybrids have identical spectra. The three hybrids from *K. aerogenes* also show identical spectra but they differ from the *K. pneumoniae* hybrids. Representatives from each class are shown in Figure 4. The spectra show the expected λgtl bands near the top of the gel and one additional band corresponding to the inserted segment of DNA. The insert bands do not correspond to any fragment of λ DNA. From the relative mobilities of the fragments, we estimate that the insert from *K. aerogenes* is about 6·7 kb and that from *K. pneumoniae* is about 9·9 kb.

Heteroduplexes between the λgtl hybrids and λimm434S7 DNAs show the expected substitution loops resulting from non-homology of the insert and of the *Eco*RI B and C fragments of λimm434S7 DNA (data not shown). From the *Eco*RI restriction spectra, it is not possible to distinguish whether the orientation of the insert DNAs in the individual representatives of the two classes is identical or reversed. Accordingly, the five *K. pneumoniae* hybrids were mixed, denatured, renatured, and mounted for electron microscopy. Only homoduplexes were found, indicating that all five hybrids were identical. All three hybrids from *K. aerogenes* were identical by the same criteria. These hybrids will now be referred to as λgtl-*Ka polA* and λgtl-*Kp polA*.

(c) *DNA polymerase I activity in extracts of hybrid-infected cells*

Cultures (1 l) of H560 (pol I mutant) and C600 (pol I wild-type) were grown to approximately 5×10^8 cells/ml. The cultures were split into four parts: uninfected, infected with λ *red*$^+$, λgtl-*Ka polA* and λgtl-*Kp polA*. After incubation for an additional 25 minutes at 37°C, DNA polymerase I levels were determined. The DNA polymerase I levels in *K. aerogenes* and *K. pneumoniae* were also measured. The data are summarized in Table 2. Both of the hybrids produce a significant amount of polymerase I activity upon infection of H560. The levels found in such infected extracts are roughly equal to those found in uninfected extracts of C600. When C600 is infected by either
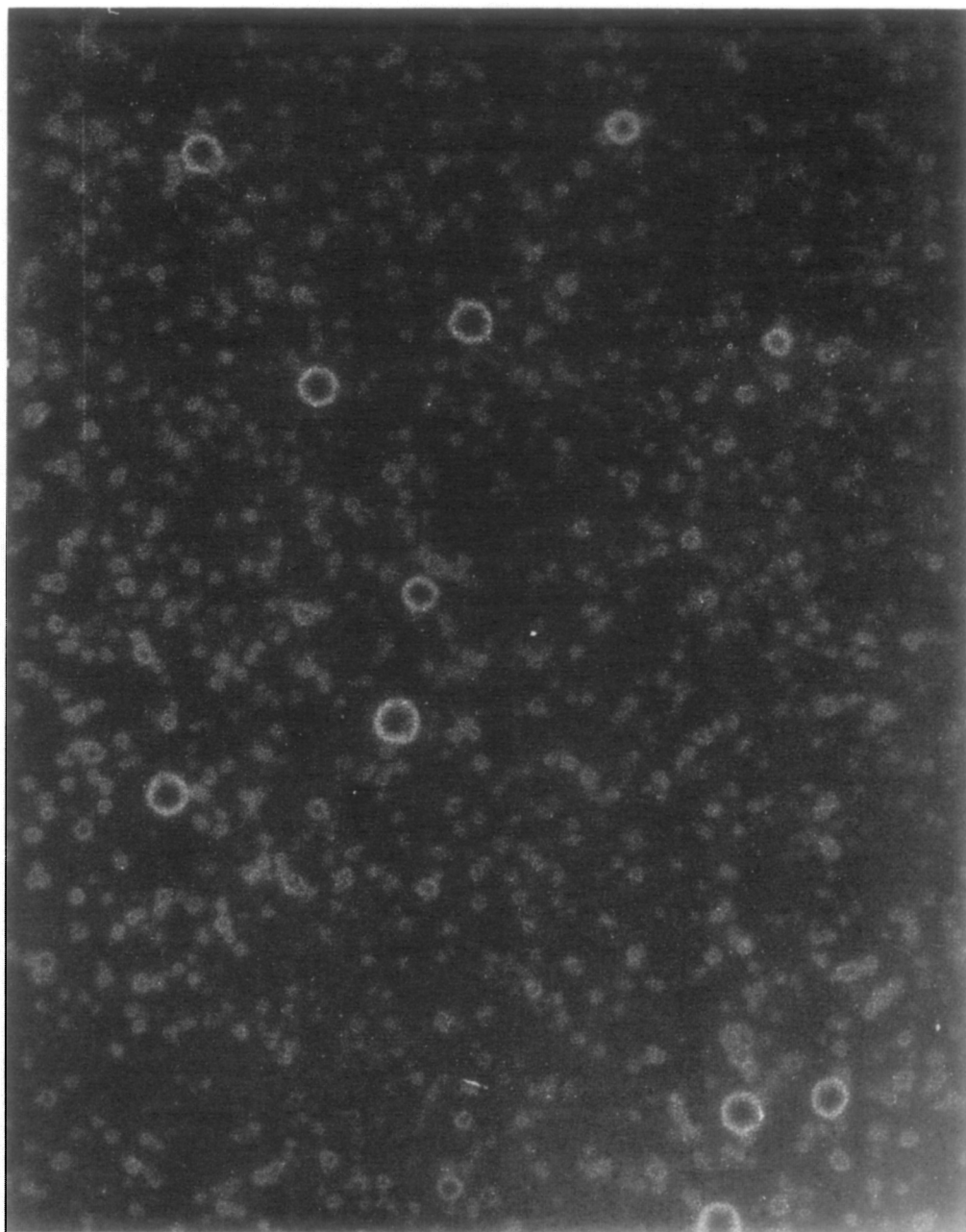
FIG. 3. Isolation of *polA*-containing phage.

10^5 phage from a λgt1-*K. pneumoniae* hybrid pool were infected into H560 (*polA*1). The plate was sprayed with 100 μg ethidium bromide/ml and plaques were visualized using an ultraviolet light source (259 nm) (Struhl & Davis, 1976).
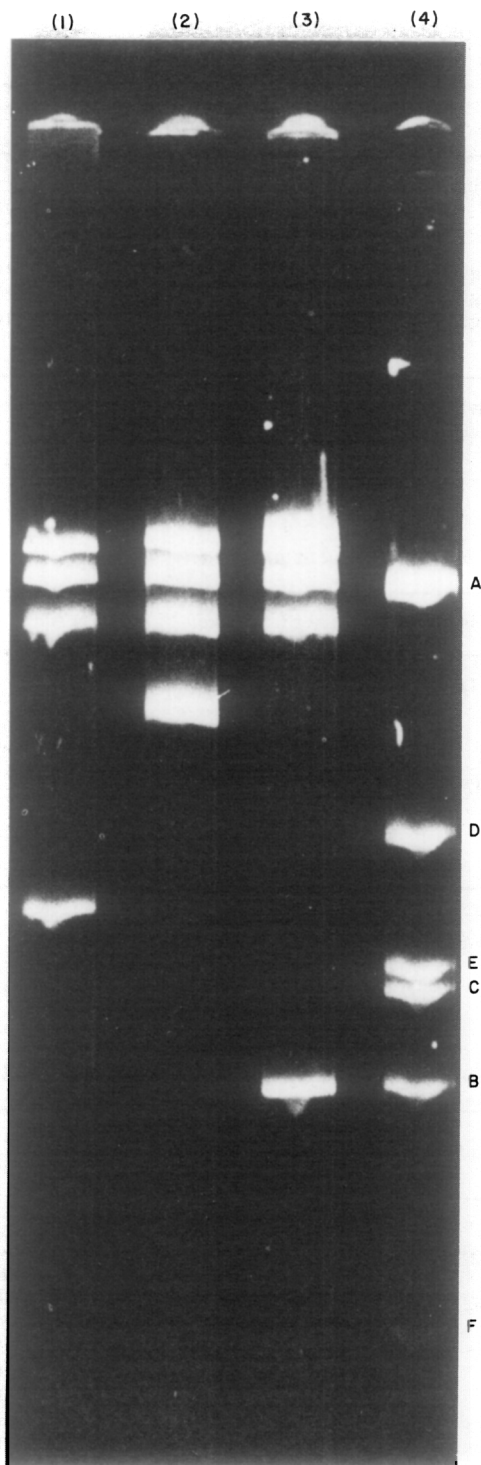
Fig. 4. EcoRI endonuclease spectra of polA-containing hybrid phages.

0·5 μg of the following phage DNAs were cleaved with EcoRI endonuclease and subjected to electrophoresis in tubes of 0·7% agarose: (1) λgt1-*Ka polA* DNA; (2) λgt1-*Kp polA* DNA; (3) λgt-λB DNA; (4) λcI857S7 DNA. EcoRI fragments are marked on the right.

TABLE 1

Efficiency of plating

Phage	C600	Bacterial host H560 (polA⁻)	ligts7
λgt-λB (red⁻)	1	<10⁻⁴	<10⁻⁴
λgt-Ka polA	1	1	<10⁻⁴
λgt-Kp polA	1	1	<10⁻⁴
λ⁺	1	1	1

Phage (listed in the left column) were titered on various hosts. Efficiency of plating on C600 ("wild-type") was defined as 1.

TABLE 2

DNA polymerase I enzymatic activity in infected extracts

Bacterial host	Infecting phage	pol I specific activity
H560 (polA⁻)	None	2
H560	λ⁺	3
H560	λgt-Ka polA	124
H560	λgt-Kp polA	98
C600 (polA⁺)	None	111
C600	λ⁺	85
C600	λgt-Ka polA	252
C600	λgt-Kp polA	207

The unit for specific activity is the number of pmol deoxy TTP incorporated into acid-precipitable cts/min per mg protein.

TABLE 3

Characterization of polymerase I enzymatic activity

Bacterial host	Infecting phage	Extract treatment		
		None	NEM	Anti-pol I
H560	None	74	0	63
H560	λ⁺	72	0	122
H560	λgt-Ka polA	4266	4845	153
H560	λgt-Kp polA	3071	2797	132
C600	None	3755	6381	71
C600	λ⁺	2838	4828	34
C600	λgt-Ka polA	7433	9591	158
C600	λgt-Kp polA	6561	8097	96
MK9000 (K. aerogenes)	None	2148	2162	74
KP1 (K. pneumoniae)	None	1632	2262	32

Data entered as cts/min. Extracts were treated with 1 mM-N-ethylmaleimide (NEM) or 5 μl of anti-E. coli DNA polymerase I antibody.

hybrid phage, the polymerase I activity is significantly higher than that of the un-infected control. The increase in polymerase activity upon infection with the hybrids is approximately equal in H560 and in C600. Mixtures of extracts result in polymerase activities which are strictly additive (data not shown). Therefore, the activities of different extracts behave independently.

DNA polymerase I activity from *E. coli* is resistant to 10 mM-*N*-ethylmaleimide and sensitive to antibody prepared against the purified enzyme. Table 3 shows that the polymerase I activity in all infected extracts as well as those of the *Klebsiella* controls are both resistant to *N*-ethylmaleimide and sensitive to the anti-polymerase I antibody.

(d) *Phenotypic expression of the polA-containing hybrids in the lysogenic state*

By the nature of their construction, all λgt1 hybrid phage are unable to form lysogens individually because they lack the genetic information for both the protein product necessary for integration (*int*) and the attachment site (*att*). Such phage can be lysogenized, however, if co-infected at 30°C with a phage capable of lysogen forma-tion (*int⁺att⁺*). The resulting double lysogens represent about 1% of the total number of lysogens (Gottesman & Yarmolinsky, 1968).

H560, by virtue of the *polA* mutation, is incapable of normal DNA repair and as a result its growth is sensitive to the mutagenic agent methylmethane sulfonate (Gross & Gross, 1969). Double lysogens of H560 with either of the two hybrids grow in the presence of 40 μg methylmethane sulfonate/ml.

(e) *Location of the* K. pneumoniae polA *gene by deletion mapping*

Wild-type λ is extremely sensitive to chelating agents and/or heat; deletion mutants derived from λ are resistant to these treatments (Parkinson & Huskey, 1971). Deletion mutants of a λgt hybrid phage containing the yeast *his3* gene were isolated in this manner. By correlating the physical structures and genetic function of these deletion mutants, the *his3* gene was mapped on the cloned DNA fragment (Struhl & Davis, 1980). Here, by analogous means, we localize the *polA* gene on the cloned *Kp polA* DNA fragment. *Kp polA* was cloned in both orientations into λgt4 DNA by the *Eco*RI–DNA ligase method. The resulting phage DNAs (λgt4-*Kp polA* and λgt4-*Kp polA'*) are equal in length to wild-type λ. Independently derived deletion mutants from each of these λgt4 hybrid phages were isolated.

The physical structures of these deletion mutants were initially analyzed by cleavage with restriction endonucleases, as described in Materials and Methods. Surprisingly, 22 out of 26 mutants derived from λgt4-*Kp polA* and 20 out of 22 mutants derived from λgt4-*Kp polA'* deleted 3·2 kb of DNA located entirely within the cloned *K. pneumoniae Eco*RI fragment. Six independent mutants from each original phage were analyzed further by *Eco*RI cleavage followed by digestion with either *Hpa*I, *Ava*I, *Ava*II or *Hinc*II endonucleases. In all cases, the 12 mutants were indistinguishable; the deletion end points in these mutants differ in map position by less than 20 nucleotide pairs. Therefore, it is likely that all 42 mutants represent independent isolates of a high frequency recombinational event. The frequency of this deletion event (0.5×10^{-4}) is about ten times higher than the sum of all *int-*

mediated deletion events (Parkinson & Huskey, 1971). Since the recombination hot spots are both localized within the *Kp polA* DNA fragment, it is not surprising that the event is independent of the orientation of the fragment in the λgt4 vector. At least one of the deletion end points occurs in the *polA* gene because none of these deletion mutants complement the *polA*1 mutation.

The high frequency deletion was mapped accurately by heteroduplex analysis. In Figure 5, a heteroduplex between λgt4-*Kp polAΔ*11 and λgt1-*Kp polA* is shown. The boundary of the *Kp polA* fragment is defined between the regions of non-homology due to the different λ vectors. Measurement of the deletion loop position localizes the Δ11 lesion between 2·4 and 5·6 kb from the left end of the cloned fragment (Fig. 7). Analysis of heteroduplex molecules between λgt4-*Kp polA'Δ*27 and λgt1-*Kp polA'* localizes the high frequency deletion to the same physical co-ordinates (data not shown).
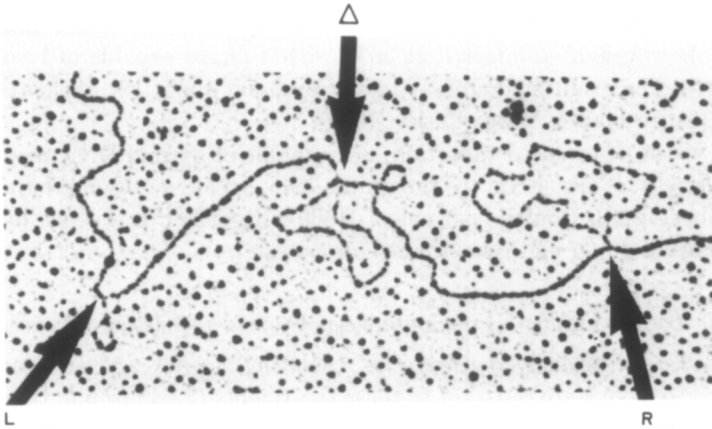

FIG. 5. Mapping the high frequency deletion mutation of *Kp polA*.
Heteroduplex between λgt4-*Kp polAΔ*11 and λgt1-*Kp polA* spread in 40% formamide according to Davis *et al.* (1971). The left (L) and right (R) end points of the *K. pneumoniae* DNA fragment are indicated.

Of the remaining six deletion mutants, four were the result of *int*-mediated recombination and were therefore mapped directly as described in Materials and Methods, section (g). All four (λgt4-*Kp polAΔ*23, λgt4-*Kp polAΔ*25, λgt4-*Kp polAΔ*26 and λgt4-*Kp polA'Δ*33) are *polA* ⁻. The results are shown in Figure 6 and the physical structures derived from them are illustrated diagrammatically in Figure 7.

Because 85% of the deletion events generate one particular lesion, mapping the *polA* gene by screening deletion mutants at random is prohibitive. To circumvent this difficulty, *polA* ⁺ deletion mutants were selected, thus eliminating the high frequency deletion event. Of 24 *polA* ⁺ deletion mutants, 16 are *int*-mediated. Because the *Kp polA* fragment is cloned in both orientations, the limits of the *polA* gene are approached from both sides. The restriction spectra of *polA* ⁺ deletions derived from either original hybrid are shown in Figure 6. The location of the *polA* gene on the
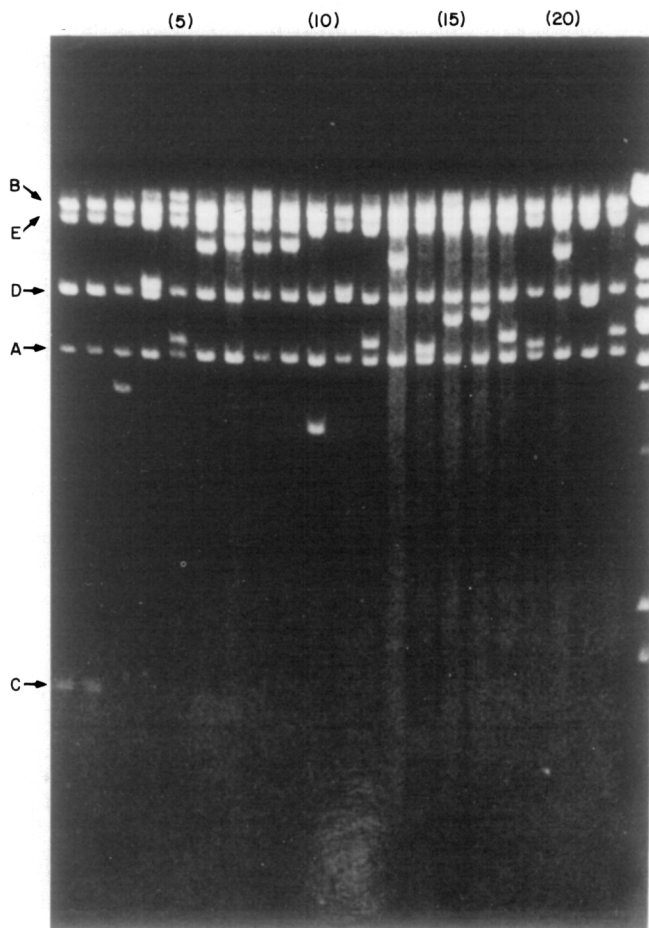
FIG. 6. Mapping deletion mutants of *Kp polA* by restriction endonuclease cleavage.
DNA rapidly prepared from crude lysates was cleaved simultaneously with *Bam*HI and *Eco*RI
and electrophoretically separated in a slab of 0·7% agarose. Samples are as follows: (1) *Δ*11;
(2) *Δ*27; (3) *Δ*23; (4) *Δ*25; (5) *Δ*26; (6) *Δ*50; (7) *Δ*51; (8) *Δ*55; (9) *Δ*57; (10) *Δ*33; (11) *Δ*59; (12)
*Δ*60; (13) *Δ*61; (14) *Δ*62; (15) *Δ*65; (16) *Δ*66; (17) *Δ*67; (18) *Δ*68; (19) *Δ*69; (20) *Δ*70; (21) *Δ*71.
Lane (22) contains λcI857S7 DNA cleaved with *Hin*dIII and *Eco*RI. Mutant DNAs in lanes (1)
and (3) to (9) are derived from λgt4-*Kp polA*, while mutant DNAs in lanes (2) and (10) to (21)
are derived from λgt4-*Kp polA*'. The mobilities of *Bam*HI-*Eco*RI fragments of λgt4 (see Fig. 1)
are indicated on the side of the Figure. All deletion mutants are *polA*⁺ except for *Δ*11, *Δ*27,
*Δ*23, *Δ*25, *Δ*26 and *Δ*33.

cloned *K. pneumoniae* DNA fragment as determined by the structures of the deletion
mutants is shown in Figure 7. The gene maps to a 3·5 kb region (at co-ordinates
4·9 to 8·4 where the units are in kb and where co-ordinate 0 is defined as the left
end of *Kp polA* when cloned in λgt4-*Kp polA*). *E. coli* DNA polymerase I has a
molecular weight of 109,000 (Jovin *et al.*, 1969); the *polA* structural gene and its
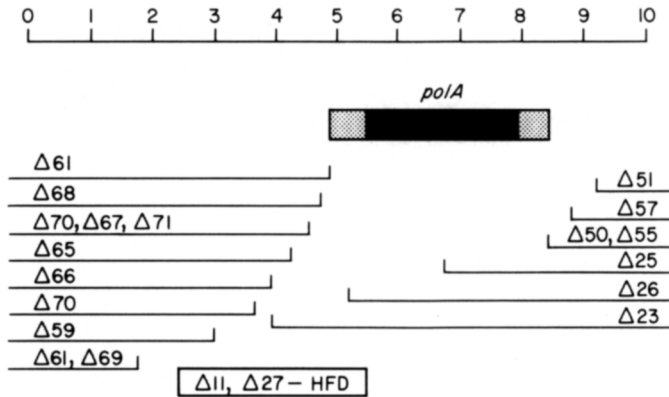promoter should be coded on at least 3 kb of DNA.

FIG. 7. Deletion map of the *K. pneumoniae polA* gene.

The co-ordinates shown on the top line are defined in kb from the left end of the *Kp polA Eco*RI DNA fragment when cloned in λgt-*Kp polA*. The limits of the location of the *polA* gene on the cloned fragment are indicated by the box. The blackened region indicates definite *polA* structural sequences; the shaded region indicates possible coding sequences. The mapping positions of the deletion mutants are shown below. HFD, high frequency deletion.

(f) *Sequence arrangement and conservation of the polymerase I genes of*
K. aerogenes, K. pneumoniae *and* E. coli

The DNA polymerase I gene regions from these enteric bacteria were directly compared at the molecular level by heteroduplex analysis of λ hybrid phages carrying *polA*. The hybrids containing the *Klebsiella polA* genes have been characterized in the previous sections; the hybrid containing the *E. coli polA* gene λ762-*Ec polA*) was characterized by and obtained from Kelley *et al.* (1977). The heteroduplex molecules are shown in Figures 8 to 9 and illustrated diagrammatically in Figure 10.

Heteroduplexes between the two different hybrids carrying the *Klebsiella polA* gene regions are shown in Figure 8. In relatively non-stringent conditions (40% (v/v) formamide), the insert regions of the molecule consist of a 5·9 kb duplex region of strong homology bounded on the left by an unpaired region and on the right by a small deletion-type loop. In 60% formamide, the small loop on the right has become a small substitution loop. In stringent conditions (70% formamide), the substitution loop on the right has become enlarged and the formerly duplex region has been shortened and divided into two parts (2·5 and 1·5 kb) by an unpaired region of approximately 0·6 kb. The substitution loop on the left is relatively unaffected by 70% formamide.

The homology pattern between the *K. aerogenes* and the *E. coli* hybrid phages strongly resembles that found between the *Klebsiella hybrids*. In 40% formamide, the *polA* DNA fragments have 4·9 kb of sequence homology (Fig. 9(a)); this represents most, but not all of the *E. coli* DNA fragment. Because the DNA fragments are cloned in different vectors, the structures at the end of the cloned fragments cannot be directly determined by simple inspection; they must be inferred by measurements of the heteroduplexes. The most likely structure is that only the right end of the *E. coli polA* fragment is not homologous to the *K. aerogenes polA* fragment. In 70%
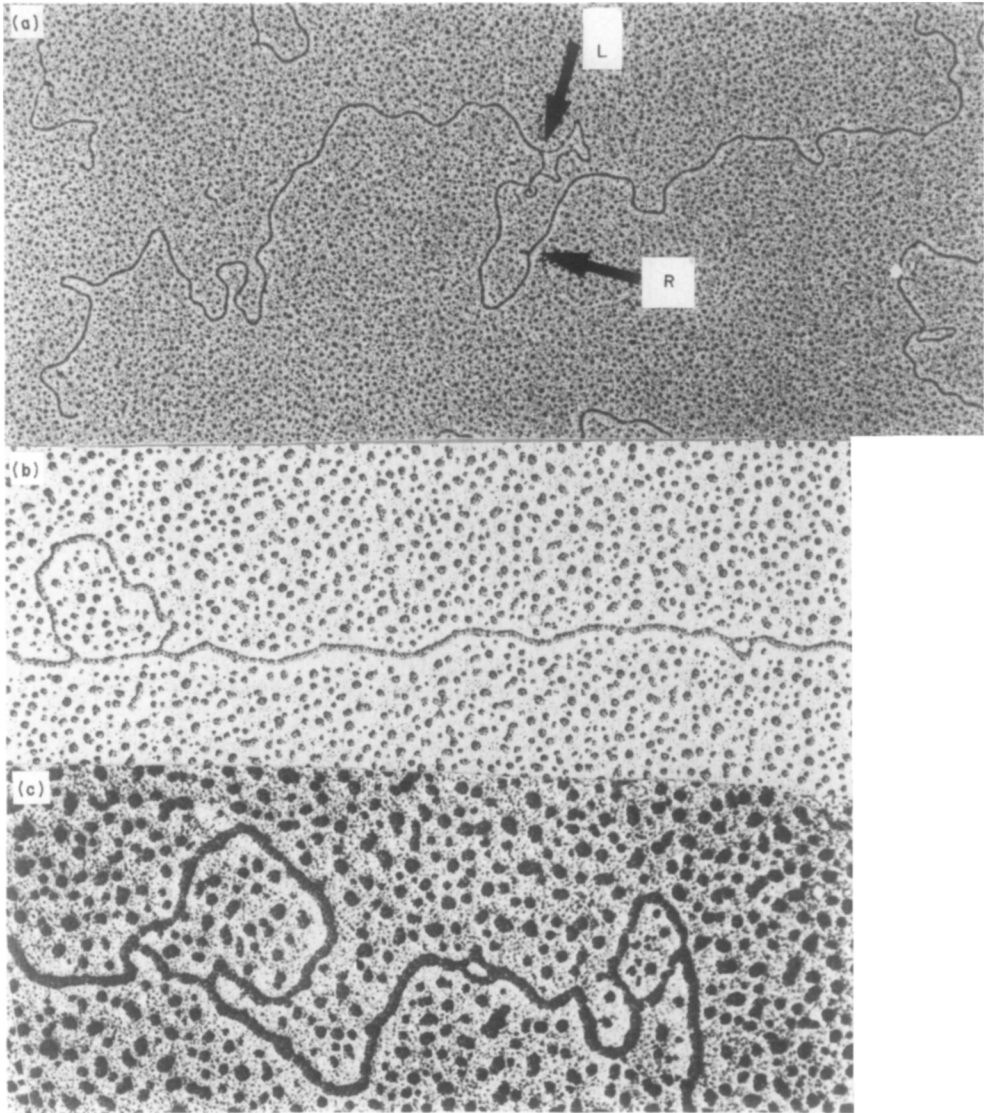
FIG. 8. Homology of the *Klebsiella polA* gene regions.
   Heteroduplex of λgt1-*Kp polA* and λgt1-*Ka polA* was spread in 40% formamide (a). The entire molecule is shown and the left and right end points of the *Klebsiella* DNA fragments are indicated. The same heteroduplex mixture was also spread in 60% formamide (b) or in 70% formamide (c). In both these cases, only the region containing the inserted *Klebsiella* DNAs is shown. The left and right ends of the insert regions are illustrated, respectively, at the left and right ends of the Figure.

formamide (Fig. 9(b)), the duplex region has been shortened and, in 75% of the molecules, divided into two parts (2·3 and 0·4 kb) by an unpaired region (0·6 kb). Under these conditions, sequences on the right are significantly melted out, while sequences on the left appear unaffected.
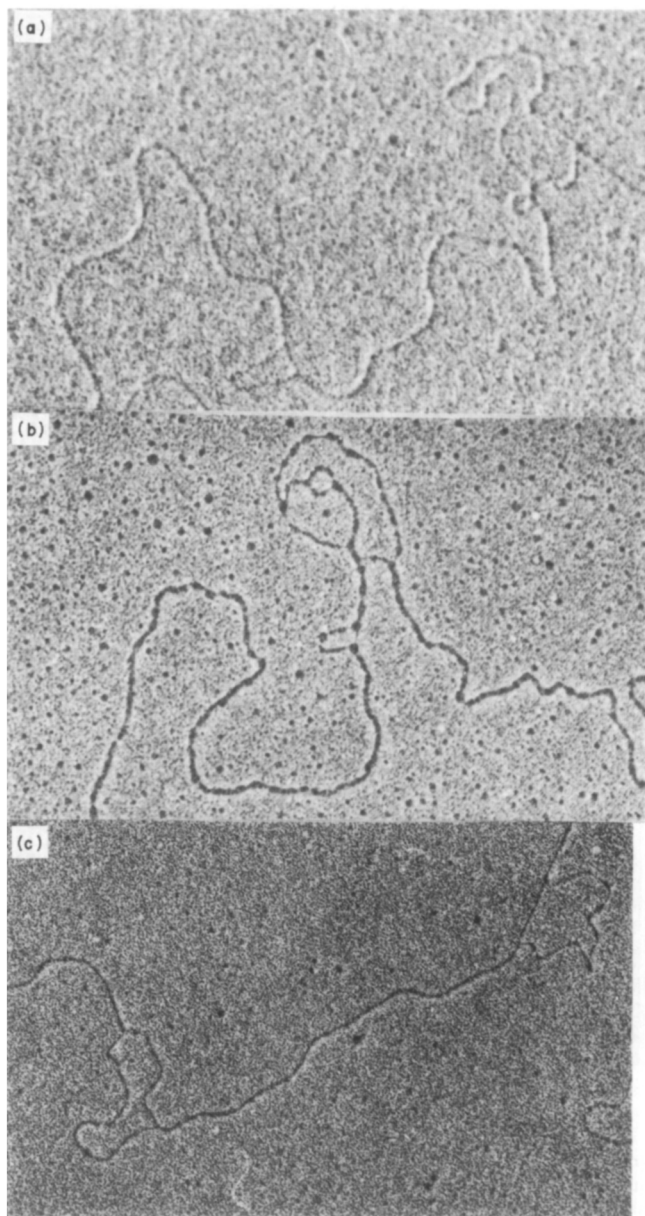
FIG. 9. Homology of the *E. coli* *polA* gene region to the *Klebsiella* regions.
Heteroduplex of λgt1-*Ka polA* and λ762-*Ec polA* was spread in 40% formamide (a) or in 70%
formamide (b). Heteroduplex of λgt1-*Kp polA* and λ762-*Ec polA* was spread in 40% formamide (c)
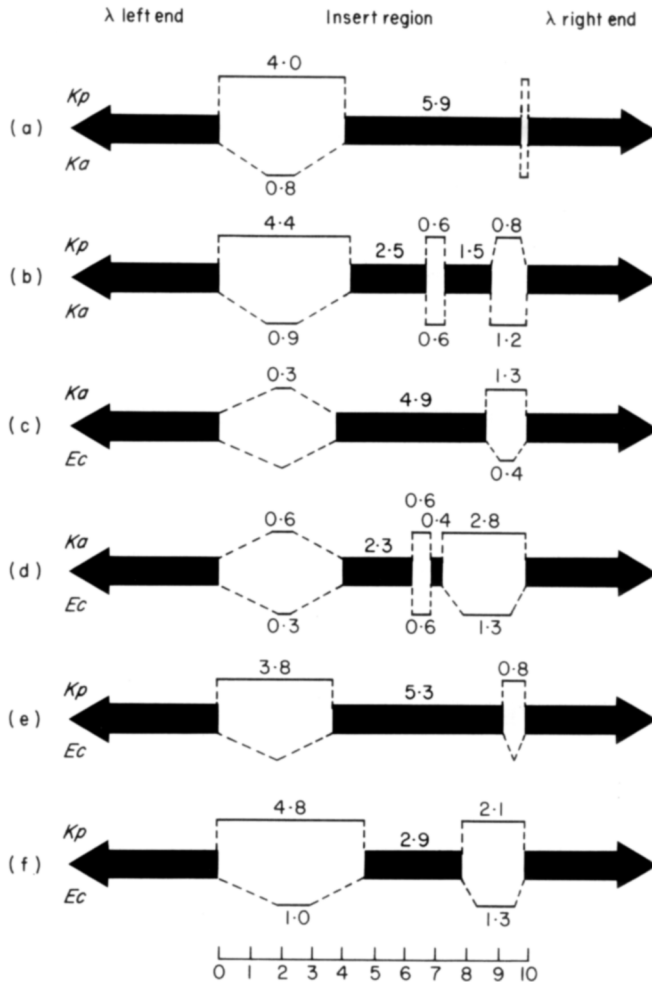or in 70% formamide (data not shown). Only the insert regions are shown.

FIG. 10. Schematic representation of enteric *polA* gene region homology.
Homologous DNA sequences are indicated by the solid bars. Lengths of DNA segments are
measured in kb. The following heteroduplex DNA molecules are shown (n = number of molecules
analyzed): (a) λgt1-*Ka polA* and λgt1-*Kp polA* in 60% formamide (n = 12); see Fig. 8(b).
(b) λgt1-*Ka polA* and λgt1-*Kp polA* in 70% formamide (n = 17); see Fig. 8(c). (c) λgt1-*Ka*
polA and λ762-*Ec polA* in 40% formamide (n = 10); see Fig. 9(a). (d) λgt1-*Ka polA* and λ762-*Ec*
polA in 70% formamide (n = 14); see Fig. 9(b). (e) λgt1-*Kp polA* and λ762-*Ec polA* in 40%
formamide (n = 10); see Fig. 9(c). (f) λgt1-*Kp polA* and λ762-*Ec polA* in 70% formamide (n = 14).

Essentially the entire *E. coli polA* fragment is homologous to the *K. pneumoniae*
polA fragment as shown by the heteroduplexes spread in 40% formamide (Fig. 9(c)).
In 70% formamide (data not shown), the duplex region is shortened to 2·9 kb and
is flanked by unpaired regions.

The results above imply that the orientation of the three inserts is the same for
the three different hybrids. By cleavage of λgt1-*Kp polA* DNA with *Eco*RI endo-
nuclease and subsequent covalent joining by DNA ligase, a hybrid was obtained

(λgt1-*Kp polA'*) which, by heteroduplex analysis, had the inserted DNA oriented in the opposite direction from that of the original isolate λgt1-*Kp polA*. This hybrid also grows on H560 but not on *ligts*7. Figure 11 shows a heteroduplex between DNAs from λgt1-*Ka polA* and the inversion λgt1-*Kp polA'*.

Surprisingly, frequent overlaps which appear to be regions of weak sequence homology are visible when spreading is done in 30% formamide. Such regions are small; the average length being less than about 200 base-pairs and the largest about 500 base-pairs. The heteroduplex on the left in Figure 11 has two such regions (A and C), and the heteroduplex on the right has three regions (A, B, C). A total of 79%
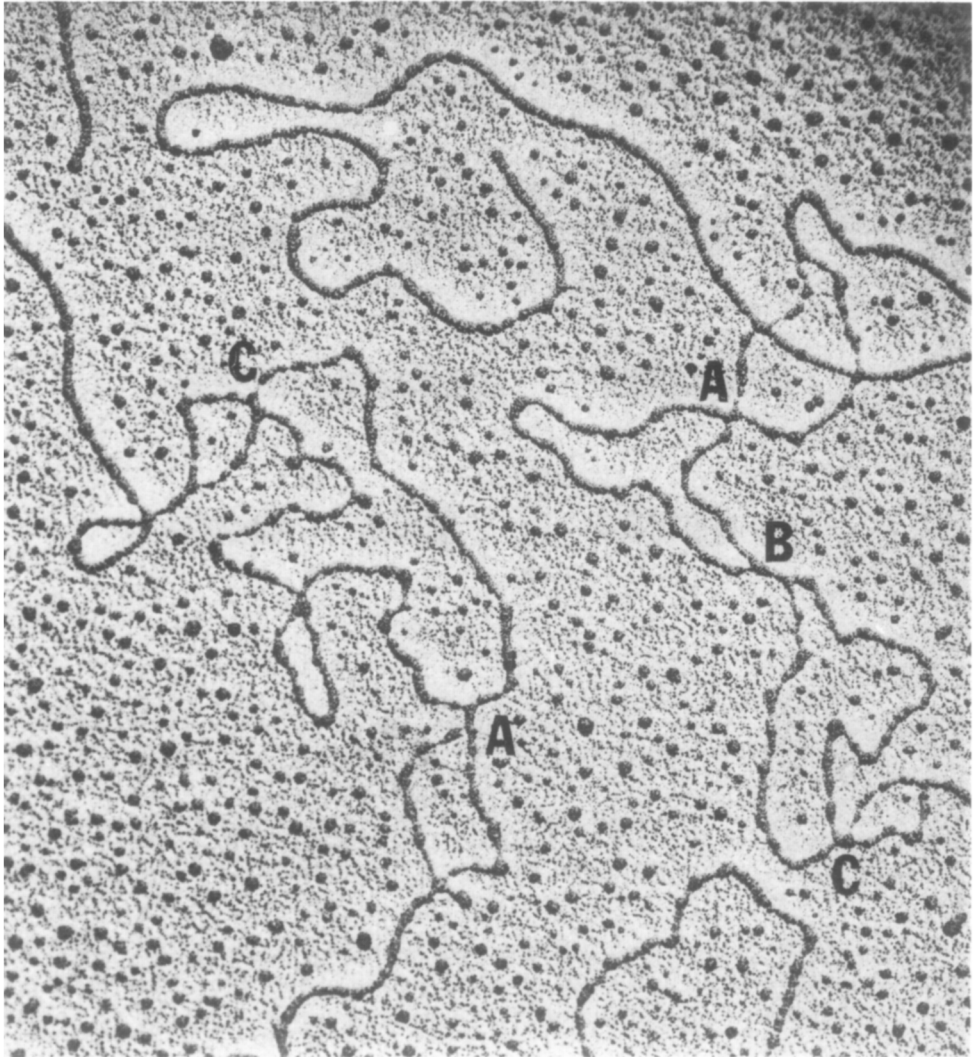

FIG. 11. Heteroduplex of λgt1-*Ka polA* and λgt1-*Kp polA'*.
Heteroduplex DNA was spread in 30% formamide. Only the inserted *Klebsiella* DNAs are shown. Weak DNA sequence homology regions A, B and C are indicated (see text).

(41 out of 52) of the heteroduplexes under these conditions have at least one overlap region. If these overlaps result from weak sequence homology, then they should occur at unique locations with respect to the total length of each insert. The data shown in Figure 12(a) indicate two significant regions of weak homology (designated A and C). Region B is found infrequently and may not be significant. Of the molecules
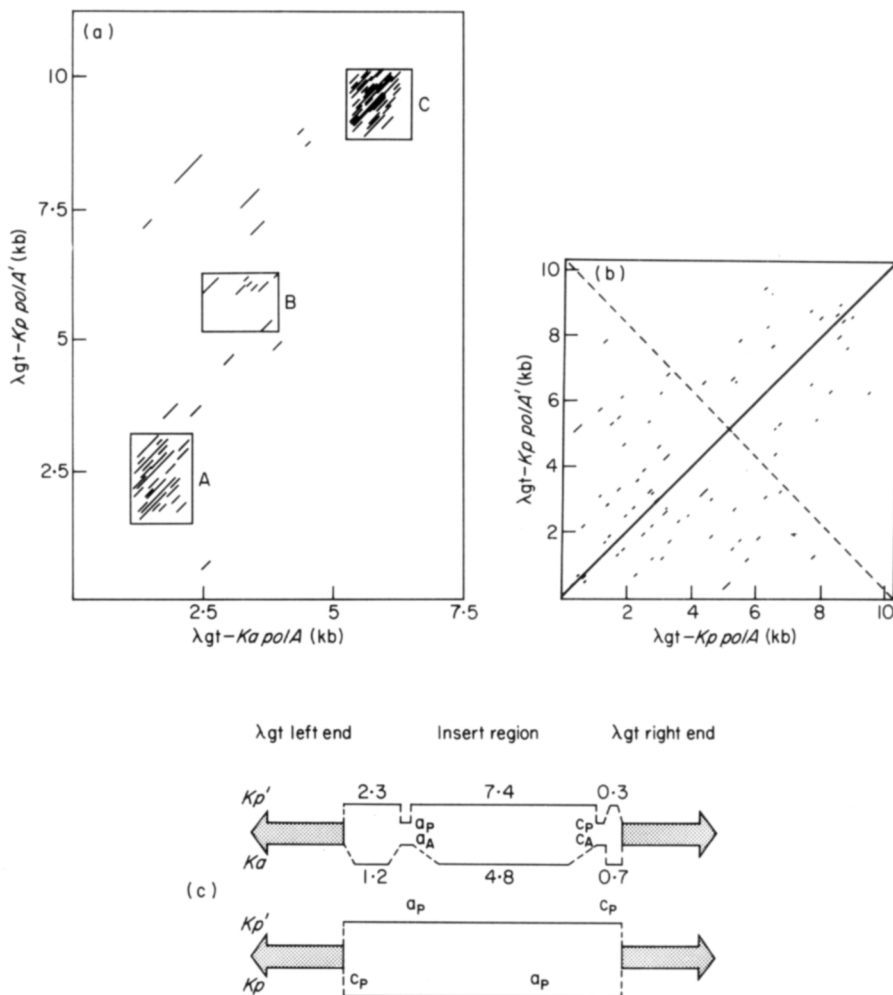

FIG. 12. Overlap maps of weak sequence homology. (a) The heteroduplex is between λgt1-*Ka polA* and λgt1-*Kp polA'*. Axes represent the length from the left end of the inserted DNA. Dashes indicate positions of apparent double-stranded regions. 55 molecules were measured and all strand crossovers are included. (b) The heteroduplex is between λgt1-*Kp polA* and λgt1-*Kp polA'*. All points (from 43 molecules) are plotted twice. Each pair is symmetric about the solid 45° line because the single strands of *Kp polA* and *Kp polA'* are indistinguishable. The broken line from the upper left to the lower right is the theoretical line for short inverted repeat sequences. A schematic representation of these heteroduplex molecules is shown in (c). The duplex λ sequences are indicated by the shaded bars. Duplex regions within the *Klebsiella* fragments are indicated by lines in close proximity, while unpaired regions are indicated by lines far apart. Sequences (such as a_A) are defined in the Discussion.
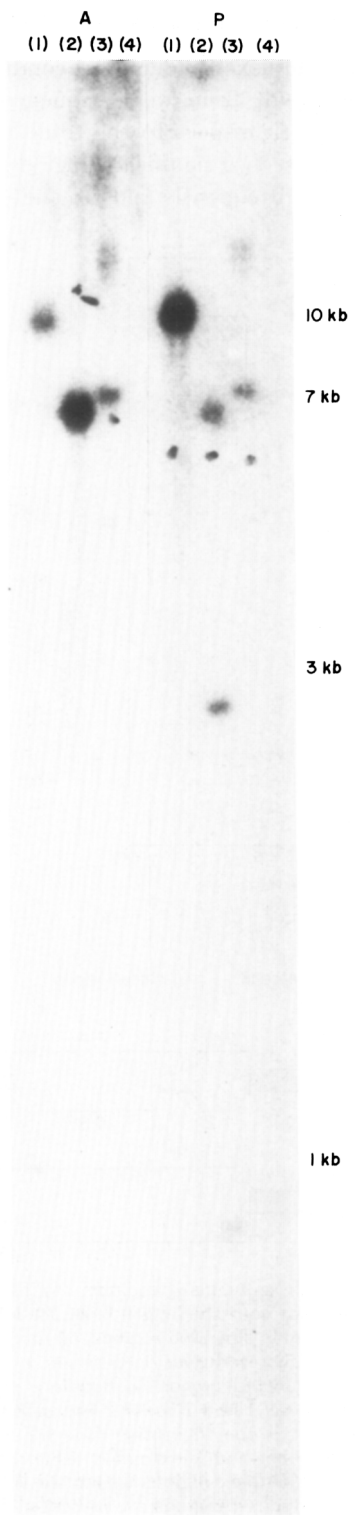
FIG. 13

that contain at least one weakly paired or overlap region, 83% are paired at A, 17% at B, and 65% at C. No unique overlap regions are seen at higher concentrations of formamide (40 to 60%). Heteroduplexes between λgtl-*Kp polA* and the inverted λgtl-*Kp polA'* DNAs in 30% formamide show only random strand overlap and no short regions of partial sequence homology can be detected (Fig. 12(b)). A diagrammatic representation of both heteroduplexes is presented in Figure 12(c).

(g) *Conservation of the polymerase I genes in other bacterial species*

In the previous section, the sequence conservation of the DNA polymerase I genes was directly analyzed with hybrid phage containing the cloned genes. In this section, sequence conservation of the polymerase I genes of *K. aerogenes*, *K. pneumoniae*, *E. coli*, *B. subtilis* and *M. xanthus* was studied indirectly by hybridization of a ^{32}P-labelled DNA probe of λgtl-*Ka polA* or λgtl-*Kp polA* across a gel of total *Eco*RI-cleaved DNA from a given organism. The autoradiograph from such an experiment is shown in Figure 13. The probe from a given hybrid is homologous to a single *Eco*RI fragment of the bacterial DNA used in the construction of that hybrid; this indicates that the cloned fragment is likely represented only once in the total genome. The length of either cloned *Eco*RI fragment is identical to the length found in the respective *Klebsiella* genome. *K. pneumoniae* has one *Eco*RI fragment homologous to the cloned *K. aerogenes* fragment; this being the result of homology of the *polA* genes. However, there are three *Eco*RI fragments of *K. aerogenes* that are homologous to the cloned *K. pneumoniae* fragment. Two of these fragments (2·8 and 0·8 kb in length) have not been cloned; presumably, they are homologous to the cloned sequences flanking the *K. pneumoniae* DNA polymerase I gene. Two *Eco*RI fragments of *E. coli* DNA (approx. 15 and 7 kb in length) are homologous to the *polA*-containing fragment from either species of *Klebsiella*. DNA from *B. subtilis* and *M. xanthus* contains no sequences detectably homologous to either radioactive probe under the hybridization conditions used.

4. Discussion

We have isolated the genes coding for DNA polymerase I from *K. aerogenes* and from *K. pneumoniae*. Viable molecular hybrids containing the polymerase I genes complement an *E. coli polA* mutation and produce functional enzyme upon infection into *E. coli* hosts.

Figure 13 indicates that the inserted *Eco*RI fragments in λgtl-*Ka polA* and λgtl-*Kp polA* indeed come, respectively, from *K. aerogenes* or *K. pneumoniae*. Since each hybrid phage contains an inserted *Eco*RI DNA fragment, which is the same length

---

FIG. 13. Autoradiograph of nucleic acid hybridization.

2 μg of DNA from (1) *K. pneumoniae*, (2) *K. aerogenes*, (3) *E. coli* and (4) *B. subtilis* were cleaved with *Eco*RI endonuclease and electrophoretically separated in a slab gel of 0·7% agarose. The DNA was denatured, neutralized, and transferred from the gel to strips of cellulose nitrate as described by Southern (1975). DNA on one of these strips was threatened with hybridization with ^{32}P-labelled nick-translated λgtl-*Ka polA* DNA (designated A). The other strip was challenged with a similar probe from λgtl-*Kp polA* (designated P). The Figure is an autoradiograph of the nucleic acid hybridization. This Figure does not include the results from hybridization of these probes to total DNA of *M. xanthus* which are mentioned in the text. The lengths of DNA fragments were calibrated using restriction enzyme fragments of λ DNA as standards (Philippsen *et al.*, 1978).

as the unique *Eco*RI fragment in the genome from which it is isolated, it is likely that the cloned *Klebsiella* sequences have not been rearranged during lytic growth of these phages in *E. coli*. From Figure 13, we infer that the polymerase I gene from *E. coli* was probably unclonable in λgt1 because the gene is cleaved by *Eco*RI and/or because one of the two *Eco*RI fragments is too large. The *E. coli polA* gene was isolated by Kelley *et al.* (1977) as a DNA fragment generated by *Hin*dIII endonuclease cleavage.

Both hybrid phage carry the *Klebsiella* promoter for *polA* expression, because polymerase activity, as assayed by MMS resistance, occurs in double lysogens under which conditions the major λ promoters are repressed. When the phages are grown lytically, the inserted regions in λgt1 hybrids can probably be transcribed from either the early λ promoter (P_L) or the late λ promoter (P'_R). The messenger RNAs produced would be "read-through" transcription products (Franklin, 1971). During lytic infections, however, it is not clear from which promoter the polymerase genes are most actively transcribed.

Having cloned *polA* genes from three closely related enteric bacteria, it is possible to examine the conservation of the DNA sequences in and around the gene. The coding region of this gene should be about 3 kb, since the *E. coli* protein has a molecular weight of 109,000 (Jovin *et al.*, 1969). We expected extensive DNA sequence conservation of this gene since it is essential (Konrad & Lehman, 1974) and codes for a protein which is conserved by enzymological and immunological criteria (Tafler *et al.*, 1973).

The *polA* gene regions from these three organisms are extensively homologous. In typical hybridization conditions (40% formamide) the *K. pneumoniae* and the *E. coli* DNA fragments seem to be completely colinear in their conservation; the single-stranded regions of the heteroduplexes are due to differences in the cloning vectors and to the fact that the cloned fragments are of different sizes. In heteroduplexes between the two other pairs of cloned *polA* fragments, most, but not all, of the sequences are colinear; the absence of colinearity is readily apparent from the substitution loops at the ends of the cloned fragments. *K. aerogenes* and *E. coli* differ in sequence at their right ends while *K. aerogenes* and *K. pneumoniae* differ at their left (and possibly their right) ends. The fact that both ends of the cloned *K. aerogenes* DNA fragment are different from the apparently colinear *E. coli* and *K. pneumoniae* fragments suggests the possibility that during evolution, most of the *K. aerogenes* fragment was inverted with respect to its flanking sequences relative to the other enteric genomes. The breakpoints for this hypothetical inversion would be at the junctions defined by the substitution loops. It is also possible that the sequences flanking the *polA* gene in *K. aerogenes* have diverged more relative to the other two organisms or represent deletion or insertion of DNA.

In stringent hybridization conditions (70% formamide), only a subset of the previously paired sequences remain as duplex structures; they map predominantly at the location of the *polA* gene. This suggests that the *polA* gene is more highly conserved than the sequences adjacent to the gene. In two out of the three heteroduplex pairs (both involving the *K. aerogenes* DNA fragment), the highly conserved regions are split by an internal bubble. The apparent size of this bubble may be an overestimate, because small regions of low or non-homology can be extended under these spreading conditions. Because length measurements of single-stranded regions

are somewhat unreliable, it is not possible to distinguish with certainty whether this internal bubble maps at the end of or within the *polA* gene.

Heteroduplex analysis gives the surprising result that there is duplex formation between a given DNA strand of the *polA*-containing *Eco*RI fragment of *K. aerogenes* and both DNA strands of the *polA*-containing *Eco*RI fragment of *K. pneumoniae*. In this discussion, it will be assumed that this duplex formation results from the weak pairing of partially homologous base sequences. The weak pairing found between identical strands of the *polA*-containing *Eco*RI fragments is localized by the method of Ferguson & Davis (1975). This method is very sensitive, since weakly homologous regions can be brought into close contact by the strong homology of the common λgt1 ends of the hybrid DNA molecules. Weak homology can be differentiated from random association of single-stranded DNA by determining the reproducibility of the contact points from molecule to molecule. Regions A, C, and possibly B, show weak but significant homology by this test (Fig. 12). The control of heteroduplexing one of the hybrid DNAs with its own inversion yielded a random distribution of contacts under the same experimental conditions. Thus, apparent weak sequence homology is probably not due to interstrand pairing of short inverted repeat sequences. The broken line in Figure 12(b) represents the theoretical predictions for the location of the double-stranded regions which would occur as a result of short inverted repeat sequences. No pairing is observed along this line. However, it is possible that the necessary structures would be unstable and therefore not observed. Although the identical strand pairing is not solely due to annealing of short inverted repeat sequences, the possibility that the weakly homologous regions may contain such sequences remains.

We can only speculate about the nature of the weak sequence homology (Fig. 14). To simplify the discussion, we define the weakly homologous sequences by the region (designated a,b,c) and by the organism (subscript A or P). The A region of Figure 11 results from the pairing of a_A and a_P. Perfect homology of these sequences in homoduplexes is designated a_A and $a_{A'}$ or a_P and $a_{P'}$. There is no assumption that $a_P = a_{A'}$ or that the pairing of $a_{A'}$ and $a_{P'}$ would behave identically in the electron microscope as a_A and a_P. However, if the observed pairing results from base sequence homology, then a sequence of *K. aerogenes* DNA is also present (to the extent defined by reproducible overlaps) in *K. pneumoniae* in the opposite orientation with respect to the *polA* gene, on the opposite strand, in an inverted orientation. It is possible that a_A, b_A and c_A share some sequence relationship but that such pairing is not observed because of constraints of the fixed interstrand DNA concentration in the heteroduplexes. Figure 14 localizes the weakly homologous sequences defined by the regions in Figure 12 with respect to the known structural information concerning the strong conservation between the cloned enteric *polA*-containing DNA fragments defined by Figures 8 to 10. Though there is considerable uncertainty in single-strand length measurements of molecules spread in 70% formamide, it is interesting that these weakly homologous sequences occur in or near regions which are not paired in 70% formamide.

The sequences defined by region B (Fig. 12) also fit this pattern. The measurements are consistent with the possibility that these weakly homologous sequences occur near the junctions of highly conserved and non-conserved sequences.
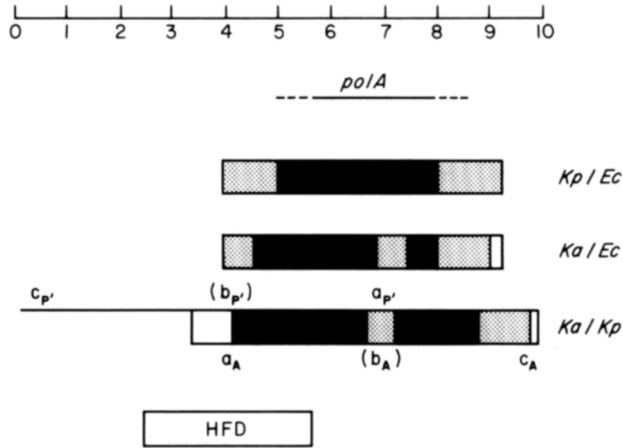
FIG. 14. The enteric *polA* gene regions.

The top line localizes the *K. pneumoniae polA* gene on the cloned DNA fragment (see Fig. 7 for details). The next 3 lines summarize the conservation data for pairs of *polA* gene regions. The solid bars indicate extremely strong homology (duplex DNA in 70% formamide). The shaded bars indicate reasonable DNA sequence homology (duplex DNA in 40% but not in 70% formamide). The open bars indicate no detectable homology. The locations of the weakly homologous regions between *Ka polA* and *Kp polA'* defined in Figs 11 and 12 are indicated on the appropriate line. The position of the high frequency deletion (HFD) event (Figs 5 and 7) is indicated on the bottom line.

Four basic classes of models are suggested in an attempt to explain the unexpected weak homologies between λgt1-*Ka polA* and λgt1-*Kp polA'* in terms of evolutionary divergence of the two *Klebsiella* species from a common ancestor. First, the weakly homologous sequences within the *K. aerogenes* and *K. pneumoniae* genomes could have been transposed in a manner similar to that described for insertion sequences (for a review, see Kleckner, 1977). This seems unlikely because separate events are required to explain each weakly homologous region. Second, a series of genetic rearrangements, with breakpoints at the sequences of weak homology, could invert the strongly conserved region(s) from one of these organisms in relation to its adjacent sequences. The weakly homologous sequences are conserved but not inverted in this model. The existence of non-conserved regions evidenced by the substitution loops in Figures 8 and 9 is consistent with this model. However, at least two rearrangements are required because the location of $a_{p'}$ is well within the 5·9 kb duplex region in Figure 14. Third, the apparent requirement for alteration of the genomic structure by illegitimate recombination can be circumvented by assuming that these short regions of weak sequence homology are non-tandemly repeated within a given genome. Short non-tandem repeats have been observed at the 5' and 3' ends of some transcribed and mobile genes of *D. melanogaster*, *S. cerevisiae*, and prokaryotic transposable elements (Finnegan *et al.*, 1978; Potter *et al.*, 1979; Cameron *et al.*, 1979; Kleckner, 1977). In this respect, it is interesting that the distance between the pair defined by a_A and $a_{p'}$ could be large enough to accommodate the polymerase I gene. The mapping position of the *polA* gene determined by deletion analysis is not inconsistent with this idea. In this case, the weakly homologous regions might

represent spacers between active genes. In other organisms, spacers are frequently very rich in dA and dT residues (Portman *et al.*, 1976; Goldberg, 1979; Struhl & Davis, unpublished results); this could account for the weak homology. In addition, it could account for the tendency of these sequences (when heteroduplexed in the same orientation) to denature in high concentrations of formamide. Fourth, the weak homology could be fortuitous, artifactual, or of opaque significance. It is conceivable, though unlikely, that these overlaps could result from specific interactions not related to nucleic acid base-pairing such as local structure of the DNA sequences or from constraints of the heteroduplex structure. No examples of this have been documented previously.

Another indication that the *K. pneumoniae polA* DNA fragment contains weakly homologous sequences comes from the fact that a particular deletion event within this fragment occurs at very high frequency. One of the deletion end points is likely to map within the *polA* gene because derivatives with this mutation do not complement the *polA* lesion of H560. It is likely that the deletion event results from recombination between two sequences of *Kp polA* DNA which are partially homologous in the direct orientation. Given that the deletion removes 3·2 kb of DNA and that the *polA* gene is approximately this size, it is conceivable that an early ancestor of *K. pneumoniae* had two tandem copies of a primordial *polA* gene which subsequently diverged; the deletion could result from recombination between the most conserved remnants. It is clear that a better understanding of all the apparent weak sequence homology between the *Klebsiella polA* gene regions will result only from DNA sequencing.

DNA sequencing of the *polA* structural genes would examine the conservation of these genes in greater detail. Conservation of the histone genes of flies and sea urchins has been examined at the nucleotide sequence level (Lifton *et al.*, 1978; Goldberg, 1979; Kedes, 1979); it would be informative to compare the degree and the nature of that homology to the conservation of the enteric bacterial *polA* genes. Because DNA polymerase I is highly conserved by enzymological and immunological criteria (Tafler *et al.*, 1973) across the closely related enteric bacteria examined in this work, it is perhaps not surprising that the genes are highly conserved at the nucleic acid level. However, DNA polymerase I from *B. subtilis* is enzymologically and somewhat antigenically similar to the *E. coli* enzyme, yet the gene shows no detectable DNA sequence homology to the genes from either *Klebsiella* species. In addition, the *hisB* genes from *Salmonella typhimurium* and *E. coli* have sequence homology, although antibody prepared against the purified *S. typhimurium* protein (obtained from Beth Cooper and John Roth) does not inactivate the *E. coli* enzyme (Struhl, unpublished data). The relationship between conservation at the protein level and conservation at the nucleic acid level depends upon the criteria; generalizations governing this relationship remain to be elucidated.

REFERENCES

Bachmann, B. J., Low, K. B. & Taylor, A. L. (1976). *Bacteriol. Rev.* **40**, 116–167.

Baron, L. S., Gemski, P. Jr, Johnson, E. M. & Wohlhieter, J. A. (1968). *Bacteriol. Rev.* **32**, 362–369.

Brenner, D. J., Fanning, G. R., Johnson, K. E., Citarella, R. V. & Falkow, S. (1969). *J. Bacteriol.* **98**, 637–650.

Cameron, J. R., Panasenko, S. M., Lehman, I. R. & Davis, R. W. (1975). *Proc. Nat. Acad. Sci., U.S.A.* **72**, 3416–3420.

Cameron, J. R., Philippsen, P. & Davis, R. W. (1977). *Nucl. Acids Res.* **4**, 1429–1448.

Cameron, J. R., Loh, E. Y. & Davis, R. W. (1979). *Cell,* **36**, 739–751.

Davis, R. W., Simon, M. N. & Davidson, N. (1971). In *Methods in Enzymology* (Grossman, L. & Moldave, K., eds), vol. 21, pp. 413–438, Academic Press, New York.

Emmons, S. W., MacCosham, V. & Baldwin, R. L. (1975). *J. Mol. Biol.* **91**, 133–146.

Ferguson, J. & Davis, R. W. (1975). *J. Mol. Biol.* **94**, 135–149.

Finnegan, D. J., Rubin, G. M., Young, M. W. & Hogness, D. S. (1978). *Cold Spring Harbor Symp. Quant. Biol.* **42**, 1053–1063.

Franklin, N. C. (1971). In *The Bacteriophage λ* (Hershey, A. D., ed.), pp. 621–638, Cold Spring Harbor Laboratory, New York.

Goldberg, M. L. (1979). Ph.D. thesis, Stanford University.

Gottesman, M. E. & Yarmolinsky, M. B. (1968). *J. Mol. Biol.* **31**, 487–505.

Gottesman, M. M., Hicks, M. L. & Gellert, M. (1973). *J. Mol. Biol.* **77**, 531–547.

Gross, J. & Gross, M. (1969). *Nature (London),* **224**, 1166–1168.

Hoess, R. H. & Landy, A. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 5437–5441.

Jovin, T. M., Englund, P. T. & Bertsch, L. L. (1969). *J. Biol. Chem.* **244**, 2996–3008.

Kedes, L. H. (1979). *Annu. Rev. Biochem.* In the press.

Kelley, W. S., Chalmers, K. & Murray, N. E. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 5632–5636.

Kleckner, N. (1977). *Cell,* **11**, 11–23.

Konrad, E. B. & Lehman, I. R. (1974). *Proc. Nat. Acad. Sci., U.S.A.* **71**, 2048–2051.

Lehman, I. R. & Chien, J. R. (1973). *J. Biol. Chem.* **248**, 7717–7722.

Lifton, R. P., Goldberg, M. L., Karp, R. W. & Hogness, D. S. (1978). *Cold Spring Harbor Symp. Quant. Biol.* **42**, 1047–1051.

Mandel, M. & Higa, A. (1970). *J. Mol. Biol.* **53**, 159–162.

Middleton, R. B. (1971). *Genetics,* **69**, 303–315.

Panasenko, S. M., Cameron, J. R., Davis, R. W. & Lehman, I. R. (1977). *Science,* **196**, 188–189.

Parkinson, J. S. & Huskey, R. J. (1971). *J. Mol. Biol.* **56**, 369–384.

Philippsen, P., Kramer, R. A. & Davis, R. W. (1978). *J. Mol. Biol.* **123**, 371–386.

Portman, R. W., Schaffner, W. & Birnstiel, M. (1976). *Nature (London),* **264**, 31–33.

Potter, S. S., Brorein, W. J., Dunsmuir, P. & Rubin, G. M. (1979). *Cell,* **17**, 415–427.

Rigby, P. W. J., Diekmann, M., Rhodes, C. & Berg, P. (1977). *J. Mol. Biol.* **113**, 237–251.

Southern, E. (1975). *J. Mol. Biol.* **98**, 503–517.

Streicher, S. L., Gurney, E. & Valentine, R. C. (1971). *Proc. Nat. Acad. Sci., U.S.A.* **68**, 1174–1177.

Streicher, S. L., Bender, R. A. & Magasanik, B. (1975). *J. Bacteriol.* **121**, 320–331.

Struhl, K. & Davis, R. W. (1976). In *Molecular Mechanisms in the Control of Gene Expression* (Nierlich, D. P., Rutter, W. J. & Fox, C. F., eds), pp. 495–506, Academic Press, New York.

Struhl, K. & Davis, R. W. (1980). *J. Mol. Biol.* **136**, 309–332.

Tafler, S. W., Setlow, P. & Levine, L. (1973). *J. Bacteriol.* **113**, 18–28.

Thomas, M. & Davis, R. W. (1975). *J. Mol. Biol.* **91**, 315–328.

Thomas, M., Cameron, J. R. & Davis, R. W. (1974). *Proc. Nat. Acad. Sci., U.S.A.* **71**, 4579–4583.

Tyler, B. & Goldberg, R. (1976). *J. Bacteriol.* **125**, 1105–1111.