

## Baboon Endogenous Virus Genome

### I. Restriction Enzyme Map of the Unintegrated DNA Genome of a Primate Retrovirus

MAURICE COHEN,<sup>1,2\*</sup> MARGERY O. NICOLSON,<sup>3</sup> ROBERT M. McALLISTER,<sup>3</sup> MAVIS SHURE,<sup>1,2</sup>  
NORMAN DAVIDSON,<sup>1</sup> NANCY RICE,<sup>2</sup> AND RAYMOND V. GILDEN<sup>2</sup>

*California Institute of Technology, Pasadena, California 91125,<sup>1</sup> Childrens Hospital of Los Angeles, Los Angeles, California 90054,<sup>3</sup> and Frederick Cancer Research Center, Frederick, Maryland 21701<sup>2</sup>*

A detailed restriction map was deduced for the genome of an endogenous retrovirus of a higher primate, that of baboon. The cleavage sites for 12 restriction enzymes were mapped. The unintegrated linear viral DNA intermediate that is produced by infection of permissive cells with baboon endogenous virus was isolated. Hybridization with a strong-stop complementary DNA probe demonstrated the presence of a terminal repetition in the linear viral DNA. The positions of restriction sites for two particular enzymes, *Sma*I and *Xho*I, near each end were consistent with this result and indicated that the length of the repetition is  $0.55 \pm 0.01$  kilobase. The linear viral DNA had a unique restriction map indicating that it is not a set of random circular permutations of the RNA genome. From hybridization with a 3'-specific probe, the DNA restriction map was aligned relative to the 5'-to-3' orientation of the viral RNA. We observed a minor heterogeneity in a *Bam*HI recognition site 1.95 kilobases from the right end of the linear map.

There have been several independent isolates of closely related or identical retroviruses endogenous in the genome of the baboon genus, *Papio*. These viruses are referred to generically as baboon endogenous virus (BaEV). The viruses are classified as xenotropic because they grow poorly on baboon cells (3, 14, 25). BaEV infection of heterologous permissive cells results in virus production and in integration of viral sequences into the host DNA. The number of baboon proviral genomes which become stably integrated is low, one to two copies for the M7 isolate in a canine thymus cell line (4) and five to six copies for the BAB8-K isolate in a human rhabdomyosarcoma cell (L. Allan and N. Rice, M. Cohen, unpublished data). In contrast, baboon chromosomal DNA contains as many as 95 proviral copies (16).

To study the sequence organization of integrated BaEV genomes, we have isolated unintegrated duplex viral DNA from freshly infected permissive cells and have characterized it by restriction endonuclease digestion experiments. We believe that the studies reported here are the first detailed restriction mapping of a primate retrovirus genome.

#### MATERIALS AND METHODS

**Isolation of viral DNA.** BAB8-K virus, an isolate of baboon endogenous virus (25), was used to infect a subclone of the human rhabdomyosarcoma line, RD

(15), in roller bottles at a multiplicity of infection of 0.5 virus particles per cell. The virus stock consisted of filtered (0.45  $\mu$ m) 12-h harvests of culture fluids from the human rhabdomyosarcoma producer cell, RD (BAB8-K). After a 11.5-h incubation at 37°C, the cells were gently lysed by addition of a solution containing 5 mM Tris-hydrochloride (pH 7.5), 10 mM EDTA, and 0.6% sodium dodecyl sulfate (SDS). Then 0.25 volume of 5 M NaCl was added and, after gentle inversion, the extract was left overnight at 3°C. The mixture was centrifuged at 12,000 rpm in a Sorvall SS-34 rotor for 40 min at 4°C (12). After phenol extraction of the supernatant fraction and dialysis, the supernatant was concentrated 25-fold with *sec*-butanol; it was then ether extracted to remove the *sec*-butanol. Covalently closed circular duplex DNA was fractionated from other supernatant DNAs by banding in a cesium chloride-propidium diiodide (CsCl-PI<sub>2</sub>) density gradient (8). Closed circular duplex pBR322 DNA (gift of Pauline Yen) was added as a density marker. DNA from the supercoil and relaxed DNA bands was removed with a pipette, and the dye was removed by extraction with *n*-butanol containing 1% Sarkosyl.

**Restriction endonuclease digestions.** Restriction enzymes used in this study were purchased from New England Biolabs. From 0.2 to 0.4 ng of unintegrated viral DNA (as estimated by comparison with the signal from a known mass of BaEV proviral DNA blotted in parallel) was digested for 1 h in 30  $\mu$ l of appropriate buffer with an excess of restriction endonuclease. The extent of digestion was usually monitored by including a marker DNA of known cleavage pattern in the digestion mixture. Alternatively, we ascertained that the resulting pattern of viral DNA

bands was unaltered by longer enzymatic digestion. Double-enzyme digests were performed simultaneously when the buffers for the two enzymes were compatible. If they were not, the enzyme requiring lower salt (or sulfhydryl compound) was added first. Then enough salt (or sulfhydryl compound) was added to achieve the second condition, and the second enzyme was added. Digestions were stopped by addition of EDTA.

**Gel electrophoresis.** Restriction enzyme digests either were added directly to vertical slab agarose gels (SeaKem) or were precipitated first with isopropanol, recovered by centrifugation, and dissolved in a small volume of low-salt buffer. The electrophoresis apparatus was purchased from Watson Products, Inc. (Altadena, Calif.). Depending upon the length of fragments to be resolved, we prepared 0.4-cm-thick slab gels of either 0.7, 1.0, or 1.5% agarose. Restriction fragments of  $\lambda$  DNA or pBR322 DNA (22) were used as size standards. The digests either were prepared first and added to the viral DNA digests before electrophoresis or were prepared by codigestion with the viral DNA.

**Blotting.** After electrophoresis, DNA fragments were denatured and transferred from agarose gels to sheets of 0.45- $\mu$ m cellulose nitrate (Schleicher & Schuell Co., Keene, N.H.) in  $10\times$  SSC (SSC = 0.15 M NaCl plus 0.015 sodium citrate) (21). After rinsing in  $3\times$  SSC, the filter was dried in air and baked in a vacuum oven at 70°C for 3 h.

**Viral RNA.** M7 virus was propagated in the human rhabdomyosarcoma line A204 and was purified by double banding in sucrose gradients as described by Benton et al. (2). Pelleted virus was resuspended and lysed by addition of SDS to 1% and was layered onto 15 to 30% sucrose gradients and centrifuged in the SW27 rotor for 6 h at 25,000 rpm. 70S RNA appears in about the middle of the gradient under these conditions. Pooled fractions were extracted with a 1:1 mixture of phenol and chloroform and then with chloroform-octanol, and were ethanol precipitated.

**Complementary DNA (cDNA) probes.** Short, calf thymus DNA fragments prepared by DNase I digestion (24) were isolated by fractionation on a column of Sephadex G-75 (E. Scolnick, personal communication).

The mixture for synthesis of  $^{32}$ P-labeled cDNA complementary to M7 virus RNA (cDNA<sub>rep</sub>) contained 0.05 M Tris-hydrochloride (pH 8.3), 0.05 M KCl, 0.008 M MgCl<sub>2</sub>, two unlabeled deoxynucleoside triphosphates (dNTP's) at 0.1 mM and the remaining two  $\alpha$ - $^{32}$ P-labeled dNTP's at 0.17 mM, 44  $\mu$ g of 70S viral RNA per ml, 350  $\mu$ g of calf DNA fragments per ml, 0.008 M dithiothreitol, (DTT), and 600 U of purified avian myeloblastosis virus reverse transcriptase per ml (24). After synthesis at 37°C for 4 h, samples were phenol extracted and separated from unincorporated  $^{32}$ P-labeled nucleotides on a 5-ml column of Sephadex G-50. RNA was hydrolyzed by alkali treatment (0.1 M NaOH, 100°C, 10 min) and then neutralized.

Strong stop cDNA<sub>5</sub> was prepared by endogenous synthesis of detergent-disrupted M7 virions. The reaction conditions were as follows: 0.5 nmol of [ $\alpha$ - $^{32}$ P]dATP (300 Ci/mmol) and [ $\alpha$ - $^{32}$ P]dCTP (300 Ci/

mmol) was dried in a 1.5-ml siliconized polypropylene tube. It was dissolved in 30  $\mu$ l of solution containing 0.1 M Tris-hydrochloride (pH 7.9), 0.005 M magnesium acetate, 0.05 M NaCl, 0.1 mM dGTP, 0.1 mM TTP, 0.01 M DTT, 0.025% Triton X-100, and 1.9 mg of M7 viral protein per ml. The reaction time was 2.5 h at 37°C. After phenol extraction, the aqueous fraction was separated from unincorporated label by Sephadex G-50 chromatography. DNA was isopropanol precipitated and, after recovery, was dissolved in 0.1 N NaOH. RNA was hydrolyzed by boiling for 10 min, and DNA was fractionated by polyacrylamide gel electrophoresis. Accumulated cDNA products were located by autoradiography and were eluted from the gel by crushing the gel in a mortar and pestle, suspending it in low-salt buffer, and removing gel by centrifugation. cDNA was recovered by isopropanol precipitation.

Polyriboadenylic acid [poly(rA)]-selected M7 virus RNA was used in preparation of cDNA<sub>3</sub>. The reaction conditions were 20  $\mu$ M [ $\alpha$ - $^{32}$ P]dATP (300 Ci/mmol), 20  $\mu$ M [ $\alpha$ - $^{32}$ P]dCTP (300 Ci/mmol), 66  $\mu$ M TTP, 66  $\mu$ M dGTP, 0.05 M Tris-hydrochloride (pH 8.3), 0.05 M KCl, 0.008 M MgCl<sub>2</sub>, 0.008 M DDT, 10  $\mu$ g of oligodeoxythymidylic acid [oligo(dT)] per ml, 3.6  $\mu$ g of poly(rA)-selected M7 RNA per ml, and 600 U of ovarian myeloblastosis virus reverse transcriptase per ml. The cDNA<sub>3</sub> was processed as in the procedure for cDNA<sub>rep</sub>.

**Hybridization.** Baked, blotted nitrocellulose filters were wet in  $3\times$  SSC and sealed in a Seal-a-meal bag (Sears Roebuck & Co.) with prehybridization mixture containing  $5\times$  SSC,  $10\times$  Denhardt medium (6), 0.1% SDS, 24  $\mu$ g of poly(rA) per ml (omitted in cDNA<sub>3</sub> hybridization experiment), 21  $\mu$ g of RD cellular RNA per ml, and 5  $\mu$ g of denatured salmon DNA per ml. After incubation at 68°C for 2 to 4 h, part of the mixture was removed and added to the  $^{32}$ P-labeled probe. This in turn was added back to the wet filter, and the bag was resealed. cDNA concentrations were in the range of 15 to 25 ng/ml. The filter was incubated with mixing at 68°C, and reactions were allowed to continue for a period of time necessary to achieve  $3\times$  C<sub>0t1/2</sub>. This usually meant an overnight incubation.

Filters were washed for about 1 h at 65°C in several changes of a solution containing  $3\times$  SSC, 0.1% SDS, and 0.1% sodium pyrophosphate. This was repeated, but with the salt lowered to  $0.4\times$  SSC and with no SDS. If background counts were judged to be unacceptably high (with a  $\beta$ -monitor), a final 20-min wash in the same buffer at 65°C was performed. Kodak X-Omat XR-5 film was used for autoradiography.

## RESULTS

It was anticipated that BaEV infection of permissive cells would result in synthesis of linear and circular forms of unintegrated viral duplex DNAs as well as integrated forms (8, 20). To isolate the unintegrated viral DNAs for mapping by restriction analysis, we used two different procedures.

In the first, DNA was isolated from a Hirt supernatant 11.5 h after infection and was frac-

tionated by centrifugation in CsCl-PI<sub>2</sub>. Upper-band DNA (linear plus nicked circular forms) was well resolved from the lower supercoiled DNA band. In the second procedure, we lysed infected cells in Triton X-100 and isolated DNA from the postnuclear supernatant.

The DNAs isolated by either procedure were analyzed by agarose gel electrophoresis, Southern blotting (21), and hybridization to a [<sup>32</sup>P]-cDNA probe representative of the BaEV genome (24). Figure 1 presents the map of restriction endonuclease sites on the linear BaEV genome for eight restriction enzymes. The data leading to these assignments and to orientation of the map relative to the RNA genome are presented below.

This method of analysis showed that DNA from the upper band of the CsCl-PI<sub>2</sub> gradient after Hirt extraction (12) contained an intense virus-specific fragment of length about 8.8 kilobases (kb) (Fig. 2, lane A). In addition, three faint bands could be seen, the largest of which corresponded in mobility to high-molecular-weight chromosomal DNA (visible in the ethidium-stained gel as intensely fluorescent). The autoradiographic band at this position was probably due to nonspecific sticking of labeled cDNA to the mass of chromosomal DNA. Such DNA-dependent background is occasionally observed in blotting experiments with highly radioactive probes. The other faint bands may have been circular viral DNA forms and were not further characterized. When the upper-band DNA from the CsCl-PI<sub>2</sub> gradient, containing the 8.8-kb linear viral DNA, was digested by the restriction endonuclease *Pst*I or by *Pvu*II (Fig. 2), two smaller viral DNA fragments were detected; in each case their sum was 8.8 kb, within experimental error. These results show that the unintegrated linear DNA consists primarily of a single population having defined termini and is not a circularly permuted or otherwise heterogeneous population.

**Bam**HI map. *Bam*HI cleaved the linear DNA four times, resulting in fragments of 2.75, 1.95, 1.85, 1.25, and 1.0 kb and a less intense fragment of 3.75 kb (Fig. 3-6). The 1.95-1.85-kb

doublet was not resolved in the experiments presented in Fig. 3 and 4. Since *Pst*I cleaved the linear DNA once, giving 6.55- and 2.25-kb fragments, but did not alter the *Bam*HI pattern in

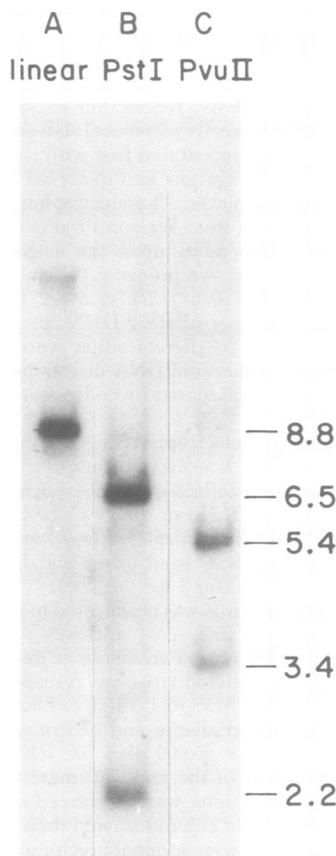


FIG. 2. Linear DNA restriction fragments. Upper-band DNA from CsCl-PI<sub>2</sub> density gradient either undigested (lane A) or digested with restriction enzymes *Pst*I (lane B) or *Pvu*II (lane C) was electrophoresed in a 0.7% agarose gel at 2 V/cm for 12 h. The DNA was transferred to a 0.45- $\mu$ m nitrocellulose sheet, baked, prehybridized, and hybridized as described in Materials and Methods. Hybridization to a cDNA<sub>rep</sub> probe (0.015  $\mu$ g/ml, 10<sup>6</sup> cpm/ $\mu$ g) was for 27 h at 68°C. The autoradiogram is a 6-day exposure.

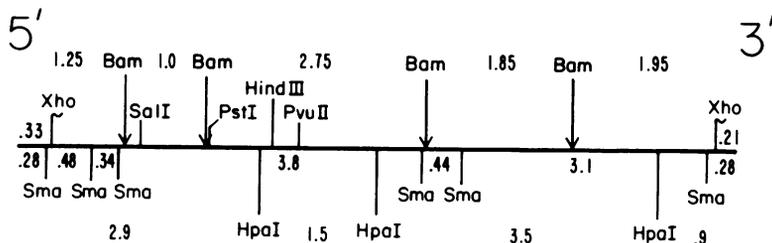


FIG. 1. Restriction map of baboon endogenous virus (BAB8-K isolate) unintegrated viral DNA.

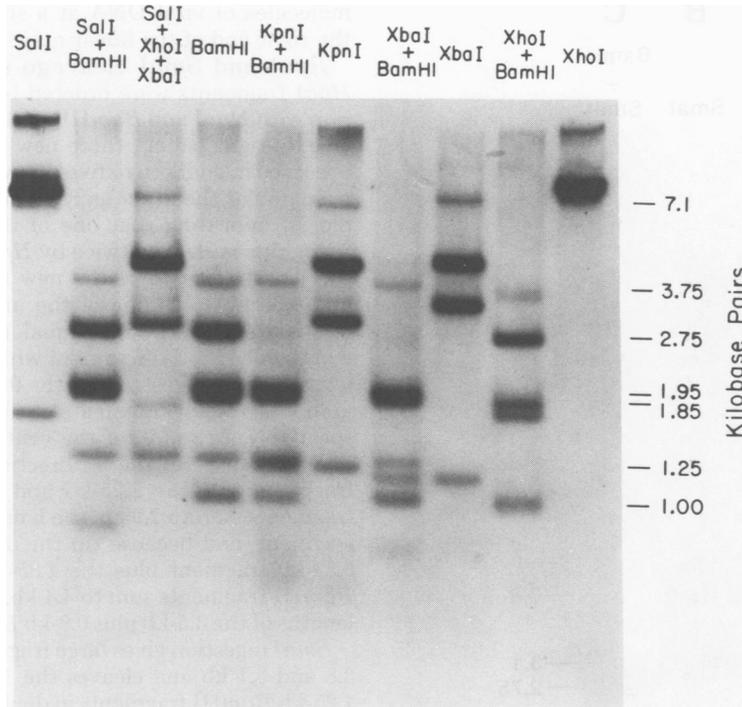


FIG. 3. Detection of unintegrated linear BaEV DNA restriction fragments. Approximately 0.2 ng of unintegrated linear BaEV DNA was digested with one or more restriction enzymes and electrophoresed in a 1% agarose gel at 1.5 V/cm for 13 h. pBR322 DNA digest length standards were electrophoresed along with the viral DNAs. They were visualized by ethidium bromide staining. Hybridization, after Southern blotting (see Materials and Methods), was with a BaEV cDNA<sub>rep</sub> probe (0.02 µg/ml,  $6 \times 10^6$  cpm/µg) for 15 h. Autoradiography was a 72-h exposure.

a double digest (data not shown), the *Pst*I site was predicted to be close to a *Bam*HI site. Therefore, the 1.25- and 1.0-kb *Bam*HI fragments were grouped on one side of the *Pst*I site and the 2.75-, 1.95-, and 1.85-kb fragments were grouped on the other.

By determining which *Bam*HI fragment is cleaved in double-enzyme digests with enzymes which cut the linear DNA once (*Hind*III and *Sal*I) or twice (*Xho*I), the five *Bam*HI fragments were unambiguously ordered (see Fig. 1). The detailed evidence for the assignments is shown in Fig. 3 and 6. As we will show later, *Xho*I is especially informative since it cuts so close to each end of the linear DNA.

It should be noted that in *Bam*HI digests the 1.25-kb 5'-end fragment often appeared less intense than the 1.0-kb fragment (Fig. 3, 6). The reason for this is not known. In double digests with *Bam*HI, *Sal*I cleavage resulted in the complete loss of the 1.0-kb fragment, thus indicating that the apparent increased intensity of this fragment with *Bam*HI alone is not the result of *Bam*HI site heterogeneity elsewhere in the ge-

nome (for instance, in the 1.25-kb fragment) such that another 1.0-kb fragment is generated.

**Heterogenous *Bam*HI site.** It was noted throughout this work that each time the linear DNA was digested with *Bam*HI, a 3.75-kb band of lower intensity appeared in addition to the five major *Bam*HI fragments. Because the length of this minor fragment was approximately equal to the sum of two different pairs of *Bam*HI major fragments (1.95 plus 1.85 kb, and 2.75 plus 1.0 kb), we considered the possibility that it was present as a result of fusion of two major *Bam*HI fragments. Over a sixfold range of the product of enzyme concentration and incubation time (at constant DNA concentration), the apparent intensity of this fragment was unchanged (data not shown). Therefore, it is unlikely that the fragment remained because of incomplete digestion. In double digests with *Bam*HI, the cleavage sites for *Kpn*I and *Xba*I were mapped (Fig. 3). *Kpn*I cut the 2.75-kb *Bam*HI fragment three times; *Xba*I cuts it twice (Table 1 and Fig. 3). In each of the double-enzyme digests with *Bam*HI shown in Fig. 3 (*Sal*I, *Kpn*I, *Xba*I, and *Xho*I), it

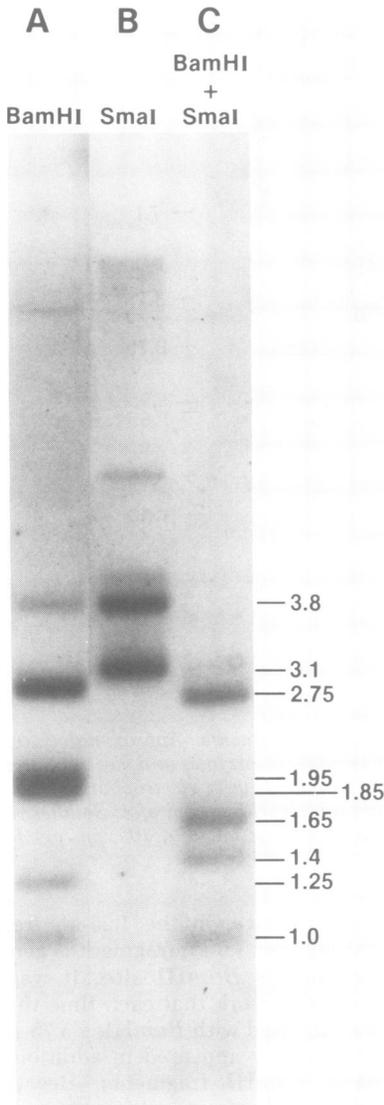


FIG. 4. Detection of BaEV linear DNA restriction fragments. About 0.2 ng of BaEV DNA from the upper  $\text{CsCl-Pf}_2$  band was digested with *Bam*HI (lane A), *Sma*I (lane B), or *Bam*HI plus *Sma*I (lane C) and electrophoresed in a 0.7% agarose gel for 12 h at 2 V/cm. Fragments were detected after blotting by hybridization with a BaEV cDNA<sub>rep</sub> probe (0.028  $\mu\text{g/ml}$ ;  $10^8$  cpm/ $\mu\text{g}$ ) for 19 h at 68°C. The autoradiographic exposure was 20 h.

was only with *Xho*I that the minor 3.75-kb fragment was reduced in size. That is, cleavage of the 1.95–1.85-kb *Bam*HI pair reduced the size of the minor band, but cleavage in the 2.75–1.0-kb *Bam*HI pair (by any of several enzymes) did not alter its size. The minor 3.75-kb band, therefore, is the result of failure by *Bam*HI to cut some

molecules of viral DNA at a site 1.95 kb from the right end of the linear map.

***Hpa*I and *Sma*I cleavage sites.** The four *Hpa*I fragments were ordered by double digestion with *Hpa*I and *Bam*HI. In the double digest (Fig. 5, right panel), three new fragments were observed in addition to five that were unchanged from one or the other enzyme digest alone (Table 2), indicating that one of the five *Bam*HI fragments is cleaved twice by *Hpa*I. The sum of the lengths of two of the new fragments (0.67 and 0.58 kb) plus one of the unchanged *Hpa*I fragments (1.5 kb) was equal to 2.75 kb, the length of a *Bam*HI fragment which disappeared in the double digest. Thus, the 0.67-kb and 0.58-kb double-digest fragments can be assigned, respectively, 5' and 3' of the central 1.5-kb fragment because (in the 5' direction) the 0.67-kb fragment plus the 1.25-kb and 1.0-kb *Bam*HI fragments sum to 2.9 kb, the length of one *Hpa*I fragment, and because (in the 3' direction) the 0.58-kb fragment plus the 1.85-kb and 1.95-kb *Bam*HI fragments sum to 4.4 kb, the sum of the lengths of the 3.5-kb plus 0.9-kb *Hpa*I fragments.

*Sma*I digestion gives large fragments of length 3.8 and 3.1 kb and cleaves the 1.95-, 1.85-, and 1.25-kb *Bam*HI fragments in double digests (Fig. 4). Even though the small *Sma*I digestion products described below were not detected in this experiment, we were able to orient the 3.8- and 3.1-kb fragments. The 3.8-kb *Sma*I fragment must approximately span the *Bam*HI 1.0- and 2.75-kb fragments since no other uncleaved segment is long enough. Thus, one *Sma*I site resides close to the junction of *Bam*HI fragments 1.25 and 1.0 kb and another close to the junction of *Bam*HI fragments 2.75 and 1.85 kb (Fig. 1).

The small *Sma*I linear DNA digestion products were resolved by electrophoresis in a 1.5% agarose gel (Fig. 7). Their lengths were 0.48, 0.44, 0.34, and 0.28 kb. Because the 0.28-kb fragment showed a greater than expected intensity (Fig. 7, lane B), it was considered to be a possible doublet (see below).

We showed above that there are two *Hpa*I sites within the 2.75-kb *Bam*HI fragment. These two sites must, therefore, lie within the 3.8-kb *Sma*I fragment. Thus, the new 2.5-kb fragment in the *Sma*I/*Hpa*I double digest (Fig. 5) had to come from *Hpa*I cleavage of the 3.1-kb *Sma*I fragment. Since we located one *Hpa*I site 0.9 kb from the right end of the map (Fig. 1), there should be a *Sma*I site 2.5 kb further to the left, giving the 2.5-kb band in the double digest. This *Sma*I site mapped about 0.4 kb to the right of the *Sma*I site we just assigned at the right end of the 3.8-kb *Sma*I fragment, and cleavage at these two sites probably resulted in the 0.44-kb *Sma*I fragment seen in Fig. 7.

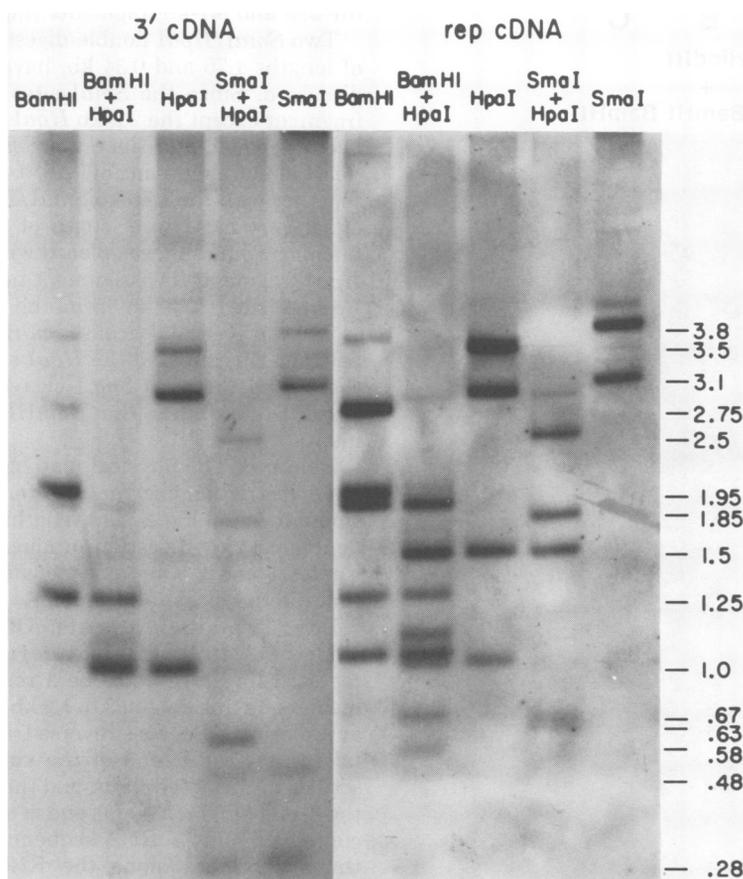


FIG. 5. Detection of BaEV linear DNA restriction fragments with BaEV representative and 3'-specific cDNA probes. Approximately 0.4 ng of linear BaEV DNA was digested with each restriction enzyme or pairs of enzymes. The digests were divided into two parts, and two lanes of each were electrophoresed in a 1% agarose gel at 2.25 V/cm for 11 h. After blotting, the filter was divided into two identical halves. The left panel was hybridized with a BaEV cDNA<sub>3</sub> probe (0.003  $\mu\text{g/ml}$ ,  $7 \times 10^6$  cpm/ $\mu\text{g}$ ) at 68°C for 11 h. The right panel was hybridized with a BaEV cDNA<sub>rep</sub> probe (0.014  $\mu\text{g/ml}$ ,  $9 \times 10^6$  cpm/ $\mu\text{g}$ ) for 26 h at 68°C. The autoradiographic exposure was 20 days. In the right panel, small restriction fragments were not detected.

The right end of the 3.1-kb *Sma*I fragment (the right-most *Sma*I site) is about 0.28 kb from the end of the genome and explains the appearance of the new 0.6-kb fragment in the *Sma*I/*Hpa*I double digest (Fig. 5) as due to *Sma*I cleavage of the 0.9-kb *Hpa*I fragment.

The length estimate of small restriction fragments seen in Fig. 7 is probably accurate to  $\pm 30$  nucleotides in the range of 1.5 to 0.3 kb. This is due to the use of pBR322 DNA restriction fragments of known sequence as size standards (22) and because of the fortuitous labeling of pBR322 restriction fragments (due to nonspecific DNA-dependent binding of probe) in the lane of the autoradiogram containing these marker fragments (Fig. 7, lane C). This allowed precise alignment of the autoradiogram with the pho-

tograph of the ethidium-stained gel. (Binding of probe to noncomplementary DNA fragments, when the DNA is present in large excess, is frequently observed in blotting experiments [T. Maniotis, personal communication].)

Of the four *Sma*I restriction fragments detected (Fig. 7, lane B), one, the 0.28-kb fragment, has already been placed at the right end of the map (see above). This fragment should be cut by *Xho*I in an *Sma*I/*Xho*I double digest to yield a 0.21-kb fragment. But we see in the double digest not only the expected 0.21-kb fragment but also a 0.28-kb fragment. Thus, the 0.33-kb *Xho*I fragment at the left end of the map (placed earlier by *Xho*I/*Bam*HI double digest) must be cut by *Sma*I at 0.28 kb from the left end. The greater than expected intensity of 0.28-kb frag-

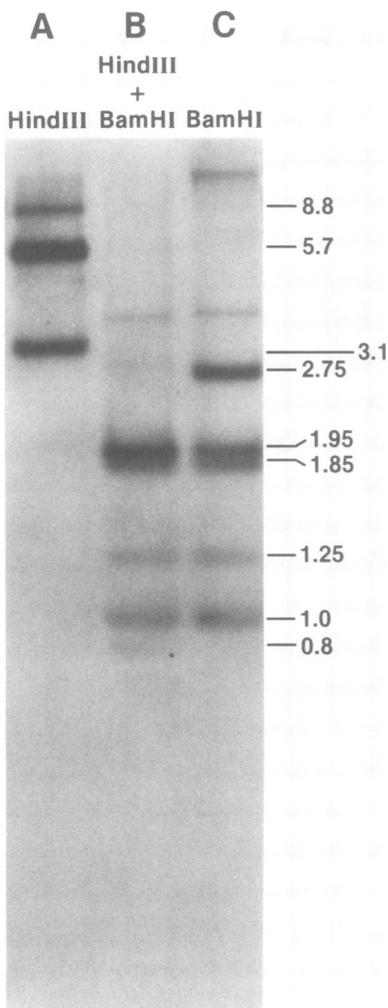


FIG. 6. Detection of linear BaEV DNA restriction fragments. Approximately 0.2 ng of DNA from the upper  $\text{CsCl-PI}_2$  band was digested with *Hind*III (lane A), *Hind*III plus *Bam*HI (lane B), or *Bam*HI (lane C). After electrophoresis in a 1.5% agarose gel at 1.5 V/cm for 13 h, DNA fragments were blotted (see Materials and Methods) and hybridized to a BaEV  $cDNA_{rep}$  probe. The hybridization mixture described in the legend to Fig. 3 was used for both filters at the same time. Autoradiography was for 72 h.

ments in the *Sma*I digest (Fig. 7, lane B) is the result, then, of two fragments of this size, one from each end of the genome.

The 0.48-kb *Sma*I fragment can be placed adjacent to the left-end 0.28-kb *Sma*I fragment because the former disappears upon cleavage with *Xho*I in the double digest and a stronger than expected 0.44-kb fragment appears (Fig. 7, lanes A and B). It was shown earlier that *Sma*I alone generates one 0.44-kb fragment between

the 3.8- and 3.1-kb fragments (Fig. 1).

Two *Sma*I/*Hpa*I double-digestion fragments, of lengths 1.76 and 0.34 kb, have not yet been discussed. Since the *Sma*I sites in each *Hpa*I fragment except the 2.9-kb *Hpa*I fragment have been assigned, and since the *Sma*I/*Hpa*I (also *Sma*I alone) fragments of 0.28, 0.48, and 0.34 kb, together with the 1.76-kb *Sma*I/*Hpa*I fragment, equal very close to a length of 2.9 kb, the remaining *Sma*I site was placed within the 2.9-kb *Hpa*I fragment. Its assignment to the left of the *Bam*HI site (1.25 kb from the 5' end of the DNA) was possible because *Sma*I did not cleave between this site and the *Hpa*I site 2.9 kb from the 5' end (because *Sma*I cleaved neither the *Bam*HI 1.0-kb nor the *Bam*HI 2.75-kb fragment).

**Resume.** The physical map of BAB8-K viral DNA restriction enzyme sites for 12 enzymes is summarized in Fig. 8. The lengths of single- and double-enzyme digestion products are given in Tables 1 and 2. One enzyme, *Pvu*I, was found that did not cleave the genome.

**Long terminal repeats.** Other retrovirus studies (13, 18, 19, 27) have revealed the unintegrated linear DNA to be a replication intermediate having long (0.3 to 1.2 kb) direct repeats at its termini. Because reverse transcription initiates near the 5' end of the viral RNA, transcribes to the 5' terminus, and then continues at the 3' end (10, 11, 23), one end of the linear DNA contains 5'-plus-3' RNA sequences. cDNA synthesis continues along the RNA to approximately the 5' end where, by an unknown mechanism, DNA complementary to both ends of the viral RNA becomes incorporated (13, 18, 19). Whether this occurs by the polymerase completing transcription at the 5' end and then translocating (a second jump) to the template's 3' end (1, 13) or by the DNA transcript incorporating a unique plus-strand DNA which contains 5'-plus-3' sequences (26), the observation is that the unintegrated linear DNA contains 5'-plus-3' RNA-specific sequences at both ends. The length of this terminal repetition appears to be specific for a particular retrovirus (5, 7, 18, 19, 27).

To prepare a probe specific for the viral RNA ends,  $^{32}\text{P}$ -labeled cDNA was synthesized in vitro upon an M7 virus RNA template by endogenous synthesis at limiting dNTP concentration. This method of synthesis is reported to result in the accumulation of several size classes of DNA including the "strong-stop" cDNA, a runoff product initiating on the endogenous tRNA primer and extending through the 5' end of the genome (10). The strong-stop cDNA in isolates of BaEV is 115 to 120 nucleotides (9).

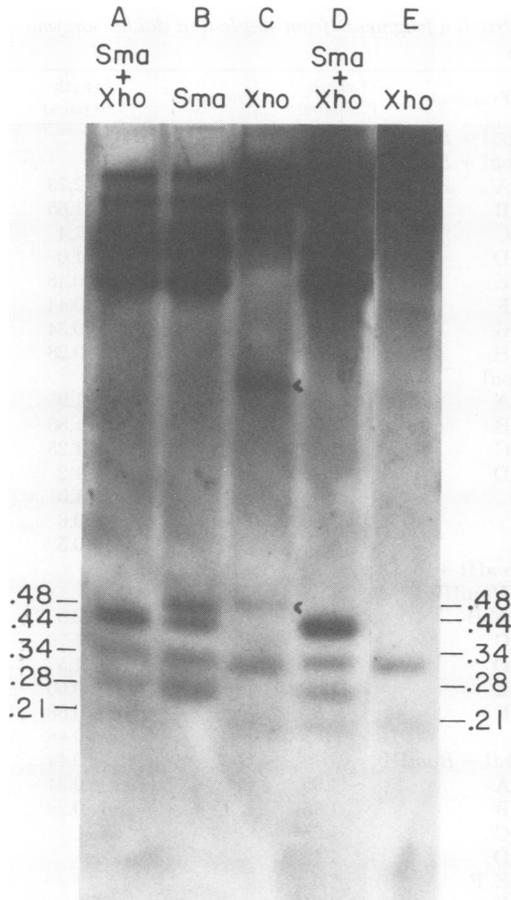


FIG. 7. Detection of small BaEV linear DNA restriction fragments. Approximately 0.5 ng of BaEV DNA from the upper  $\text{CsCl-PI}_2$  gradient band was electrophoresed in a 1.5% agarose gel at 3 V/cm for 7 h. After transfer to a nitrocellulose sheet (see Materials and Methods), the DNA was hybridized with a BaEV cDNA<sub>rep</sub> probe (0.02  $\mu\text{g/ml}$ ,  $5 \times 10^8$  cpm/ $\mu\text{g}$ ) for 16 h at 68°C. The two marked bands in lane C are the result of nonspecific binding of labeled probe to *Hin*I digestion fragments of pBR322 DNA that were electrophoresed with the experimental sample in that lane as length standards. These two *Hin*I fragments are 1.6 and 0.51 kb (21).

After alkaline hydrolysis of the tRNA primer, the cDNA products were size fractionated by polyacrylamide gel electrophoresis (Fig. 9). Strong-stop cDNA was isolated from the gel and used as a hybridization probe to blotted BaEV linear DNA digests (Fig. 10). In the case of the *Bam*HI and *Hpa*I digests, the two end fragments (1.25 and 1.95 kb for *Bam*HI, 2.9 and 0.9 kb for *Hpa*I; see Fig. 1) hybridized to the strong-stop probe. This result shows that the strong-stop sequences occur at both ends of the DNA, as

expected from the terminal repetition hypothesis. In single or double digests with *Sma*I, the penultimate 0.48-kb fragment from the left end (see Fig. 1) and a 0.28-kb fragment (right and/or left end, see Fig. 1) hybridized to the strong-stop cDNA. In the following sections, we present evidence that the long terminal repeat has a length of  $550 \pm 10$  nucleotide pairs and that the 5'-to-3' orientation of the viral RNA genome is the left-to-right orientation of the DNA map of Fig. 1. It then follows that, reading from left to right, the terminal repetition at the left end consists of 420 to 440 nucleotides from the 3' end of the viral RNA genome followed by 120 nucleotides of strong-stop cDNA, whereas at the right end, the strong-stop sequence is at the very end. This interpretation predicts that the strong-stop sequence occurs within the 0.28-kb *Sma*I fragment from the right end and within the 0.48-kb *Sma*I fragment close to the left end. These predictions are in agreement with the results in Fig. 10. Hybridization to identical digests with an approximately 50-nucleotide-long short-stop cDNA isolated from the same gel (Fig. 9) gives an identical pattern of labeling (unpublished data).

BaEV appears to replicate similar to other retroviruses since endogenous synthesis begins near the RNA's 5' end with synthesis of a strong-stop sequence (9). Thus, labeling of both ends of the linear DNA by hybridization with strong-stop cDNA is compelling evidence for the presence of a terminal repeat.

Further, we observe *Sma*I and *Xho*I sites in the same order and separated from each other by  $60 \pm 10$  nucleotides, with the *Xho*I sites 330 nucleotides from the left end and 210 nucleotides from the right end. In general, if there are sites for two enzymes at positions  $L_1$  and  $L_2$  from the left end and at positions T- $L_1$ , and T- $L_2$  from the right end, we would take this as evidence for a terminal repetition of length  $T$ . What is the probability that this occurrence is due to random chance, and not due to a terminal repetition? Let  $P$  be the probability per nucleotide of occurrence of one enzyme site (approximately the same for *Sma*I and *Xho*I). The probability that the two sites would occur by chance with a spacing of  $L_1-L_2$  ( $\pm\delta$ ) within a distance  $T$  of one end and that neither site occurs a second time within this distance is approximately  $(2\delta)TP^2(1-P)^2T$ . This is the probability of random coincidence for occurrence of equally spaced sites at both ends. We take  $P \approx 2 \times 10^{-4}$ ,  $\delta = 10$ , and  $T = 1,200$  (an estimate of the upper limit for the length of a terminal repetition based on other examples), and calculate  $5.9 \times 10^{-4}$ . Thus, the identical configuration of the *Sma*I

TABLE 1. Sizes of BaEV unintegrated linear DNA restriction fragments from single- and double-enzyme digests<sup>a</sup>

Fragment	Length (kb pairs)	Fragment	Length (kb pairs)	Fragment	Length (kb pairs)	Fragment	Length (kb pairs)
<i>Xba</i> I		<i>Xho</i> I		<i>Hpa</i> I + <i>Bam</i> HI		<i>Sma</i> I + <i>Bam</i> HI	
A	4.45	A	8.25	<i>Hpa</i> I + <i>Bam</i> HI		<i>Sma</i> I + <i>Bam</i> HI	
B	3.25	B	0.33	A	1.85	A	2.73
C	1.1	C	0.21	B	1.5	B	1.65
<i>Kpn</i> I		<i>Bam</i> HI		C	1.25	C	1.4
A	4.25	<i>Bam</i> HI + <i>Pst</i> I		D	1.05	D	1.0
B	2.85	A	2.75	E	1.0	E	0.48
C	1.2	B	1.95	F	0.9	F	0.44
D	0.5	C	1.85	G	0.67	G	0.34
		D	1.25	H	0.58	H, I	0.28
		E	1.0	<i>Hpa</i> I		<i>Kpn</i> I + <i>Bam</i> HI	
<i>Sa</i> II + <i>Bam</i> HI		<i>Xho</i> I + <i>Sma</i> I		A	3.5	A	1.95
A	2.75	A	3.8	B	2.9	B	1.85
B	1.95	B	3.1	C	1.5	C	1.25
C	1.85	C, D	0.44	D	0.9	D	1.2
D	1.25	E	0.34			E	1.0
E	0.75	F	0.28			F	0.6
F	0.25	G	0.21			G, H	0.5
<i>Sma</i> I		<i>Xho</i> I + <i>Bam</i> HI		<i>Bam</i> HI + <i>Hind</i> III		<i>Sma</i> I + <i>Hpa</i> I	
A	3.8	A	2.75	A, B	1.95	A	2.5
B	3.1	B	1.85	C	1.85	B	1.76
C	0.48	C	1.74	D	1.25	C	1.5
D	0.44	D	1.0	E	1.0	D	0.60
E	0.34	E	0.92	F	0.8	E	0.58
F, G	0.28	F	0.33			F	0.48
		G	0.21	<i>Xba</i> I + <i>Bam</i> HI		G	0.44
				A	1.95	H	0.34
				B	1.85	I, J	0.28
				C	1.25		
				D	1.1		
				E, F	1.0		
				G	0.65		

<sup>a</sup> Use of a sequenced plasmid (pBR322) DNA for length standards (22) provided restriction data reproducible to greater than ± 0.01 accuracy and justifies some length estimates to three significant figures.

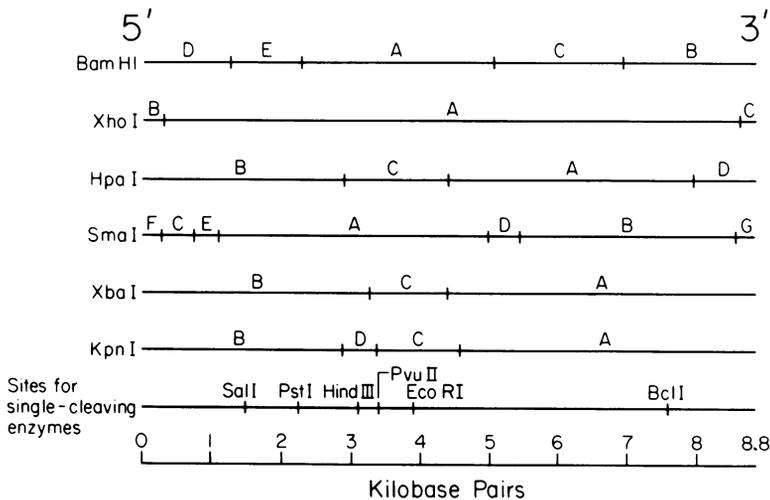


FIG. 8. Cleavage sites on BaEV (BAB8-K isolate) unintegrated linear DNA.

TABLE 2. Sizes of *BaEV* unintegrated linear DNA restriction fragments for enzymes which cleave the genome once

Fragment	Length (kb pairs)
<i>Pst</i> I	
A	6.55
B	2.25
<i>Hind</i> III	
A	5.7
B	3.1
<i>Sal</i> I	
A	7.3
B	1.5
<i>Pvu</i> II	
A	5.4
B	3.4
<i>Eco</i> RI	
A	4.9
B	3.9
<i>Bcl</i> I	
A	7.6
B	1.2

and *Xho*I sites at the two ends is strong evidence for a terminal repetition, and the calculated repetition length is  $550 \pm 10$  nucleotides.

**Orientation of the RNA genome.** The experiment described above established the presence of a terminal repetition. We then wished to determine the orientation of the RNA genome in the linear map since our original designation of 1.25- and 1.95-kb *Bam*HI fragments at the left and right end, respectively, was an arbitrary choice. That orientation as summarized in Fig. 1 derives from the following experiment.

A [ $^{32}$ P]cDNA<sub>3</sub> probe which was selectively representative of the 3' end of the viral genome was prepared by oligo(dT)-primed synthesis of cDNA from poly(rA)<sup>+</sup> viral RNA. This probe was hybridized to a blot of *BaEV* linear DNA digests (Fig. 5, left panel). Because of the terminal redundancy, intense bands are indicative of hybridization to end restriction fragments, but the relative intensity of other bands should reveal fragments progressing in from the 3' end of the RNA. Thus, in the *Sma*I digest (Fig. 5, left panel), the 0.28- and 0.48-kb end fragments were most intense (they represented a small molefraction of the genome); similarly, in the *Hpa*I digest, the 2.9- and 0.9-kb end fragments were the most intense. What is revealing, however, is which of the remaining fragments in the two digests was more intense (for comparison, hybridization of identical digests with a representative cDNA probe is shown in Fig. 5, right panel). With *Hpa*I it was the 3.5-kb and not the 1.5-kb fragment; with *Sma*I it was the 3.1-kb and not the 3.8-kb fragment. That is, fragments adjacent to the right end of the map rather than those at the left end were detected with a probe

synthesized from the RNA's 3' end. Similarly, in the *Sma*I/*Hpa*I double digest, the new intense 0.6-kb fragment was due to *Sma*I cleavage of the intense 0.9-kb fragment near the right end of the map and the new *Bam*HI/*Hpa*I 1.05-kb fragment was due to *Hpa*I cleavage of the 1.95-kb *Bam*HI fragment. Thus, the orientation 5'—RNA genome—3' obtains.

## DISCUSSION

In summary, we have observed that a free linear viral DNA intermediate is formed when *BaEV* infects sensitive cells, and have mapped a number of restriction sites on this in vivo

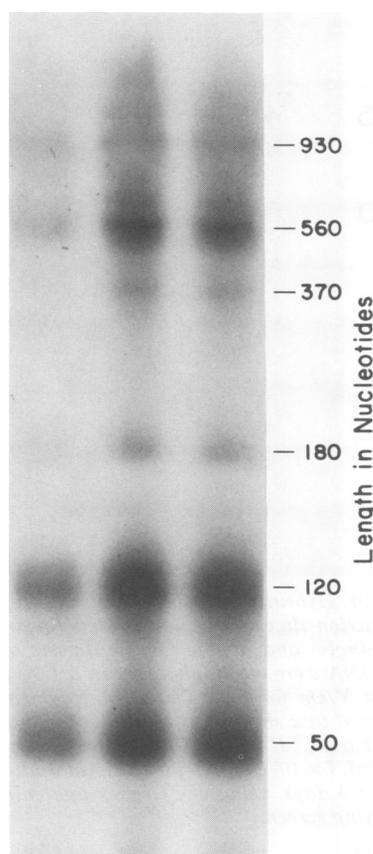


FIG. 9. *BaEV* structural stop cDNA's. At limiting dNTP,  $^{32}$ P-labeled cDNA was prepared by endogenous synthesis in detergent-disrupted M7 virions. The products were treated with alkali to destroy the tRNA primer and were electrophoresed in an 8% polyacrylamide gel for 3.5 h at 8 V/cm. The autoradiographic exposure was 12 h. cDNA was eluted from the gel (see Materials and Methods), and the fragments were sized by coelectrophoresis in an 8% polyacrylamide-7 M urea gel, using *Hae*III and *Hind*III digests of PM2 DNA as length standards.

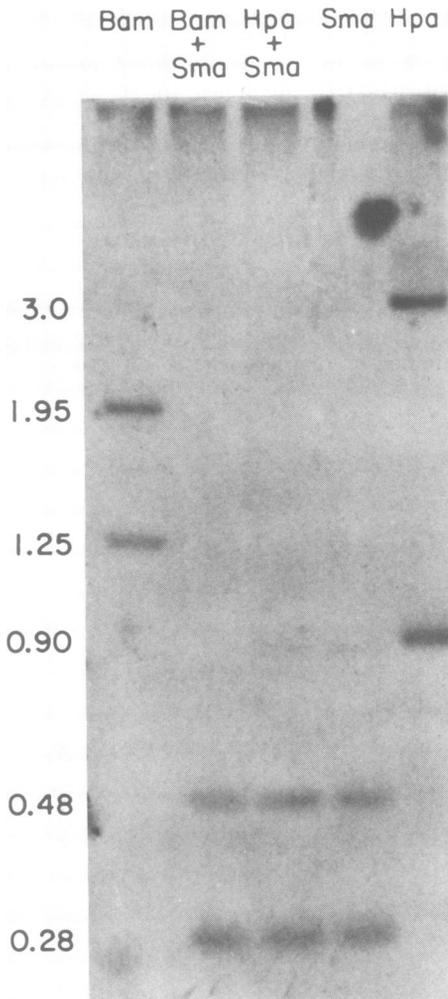


FIG. 10. Hybridization of BaEV strong-stop cDNA to restriction digests of unintegrated linear BaEV DNA. Single- and double-enzyme digests of linear BaEV DNA were electrophoresed in a 1.5% agarose gel at 4 V/cm for 8 h. The DNA fragments were transferred to a sheet of nitrocellulose and were hybridized with  $^{32}\text{P}$ -labeled BaEV strong-stop cDNA (0.5 ng/ml,  $7 \times 10^8$  cpm/ $\mu\text{g}$ ) for 15 h. Autoradiography was for 8.5 days, using one Cronex lightening-plus-intensifying screen.

DNA. The map is unique and, apart from the terminal repetition, appears to be colinear with the viral RNA. We deduce that the terminal repetition has a length of  $550 \pm 10$  nucleotide pairs. If the ends are arranged like other retrovirus genomes, they will have the form 3'5'—RNA genome—3'5'.

Because of its availability, the M7 virus isolate of BaEV was used for preparing our cDNA probes, whereas the in vivo DNA that was

mapped was the BaEV isolate, BAB8-K. Both viruses are derived from the baboon species *Papio cynocephalus*, and our hybridization results indicate that the two viral sequences are closely related and cross-hybridize readily. This is consistent with previous observations that BaEV genomes are all closely related, even among more divergent baboon species (4, 9).

This study provides the first detailed mapping data on the genome of a retrovirus of a higher primate. The results for the BaEV linear DNA are generally similar to those reported for other retroviruses.

These mapping data will be useful in our ongoing studies of the organization of the endogenous viral sequences in baboon cells and of the exogenous BaEV sequences in productively infected cells.

#### ACKNOWLEDGMENTS

This research was supported by Public Health Service contract NO1 CP 71004 from the Division of Cancer Cause and Prevention, National Cancer Institute, by Public Health Service grant F32 CA05500 from the same Institute, and by Frederick Cancer Research Center.

#### LITERATURE CITED

1. Baltimore, D., E. Gilboa, E. Rothenberg, and F. Yoshimura. 1978. Production of a discrete infectious double-stranded DNA by reverse transcription in virions of Moloney murine leukemia virus. Cold Spring Harbor Symp. Quant. Biol. 43:869-874.
2. Benton, C. V., H. M. Hodge, and D. L. Fine. 1978. Comparative large-scale propagation of retroviruses from old world (Mason-Pfizer monkey virus) and new world (squirrel monkey virus) primates. In Vitro 14: 192-199.
3. Benveniste, R. E., M. M. Lieber, D. M. Livingston, C. J. Sherr, G. J. Todaro, and S. S. Kalter. 1974. Infectious C-type virus isolated from a baboon placenta. Nature (London) 248:17-20.
4. Benveniste, R. E., and G. J. Todaro. 1974. Multiple divergent copies of endogenous C-type virogenes in mammalian cells. Nature (London) 252:170-173.
5. Benz, E. W., and D. Dina. 1979. Moloney murine sarcoma virions synthesize full-genome-length double-stranded DNA in vitro. Proc. Natl. Acad. Sci. U.S.A. 76:3294-3298.
6. Denhardt, D. T. 1966. A membrane-filter technique for the detection of complementary DNA. Biochem. Biophys. Res. Commun. 23:641-646.
7. Gilboa, E., S. Goff, A. Shields, F. Yoshimura, S. Mitra, and D. Baltimore. 1979. In vitro synthesis of a 9kbp terminally redundant DNA carrying the infectivity of Moloney murine leukemia virus. Cell 16:863-874.
8. Guntaka, R. V., O. C. Richards, P. R. Shank, H.-J. Kung, N. Davidson, E. Frisch, J. M. Bishop, and H. E. Varmus. 1976. Covalently closed circular DNA of avian sarcoma virus: purification from nuclei of infected quail tumor cells and measurement by electron microscopy and gel electrophoresis. J. Mol. Biol. 106: 337-357.
9. Haseltine, W. A., and D. G. Kleid. 1978. A method for classification of 5' termini of retroviruses. Nature (London) 273:358-364.
10. Haseltine, W. A., D. G. Kleid, A. Panet, E. Rothen-

- berg, and D. Baltimore. 1976. Ordered transcription of RNA tumor virus genomes. *J. Mol. Biol.* **106**:109-131.
11. Haseltine, W. A., A. M. Maxam, and W. Gilbert. 1977. Rous sarcoma virus is terminally redundant: the 5' sequence. *Proc. Natl. Acad. Sci. U.S.A.* **74**:989-993.
  12. Hirt, B. 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. *J. Mol. Biol.* **26**:365-369.
  13. Hsu, T. W., J. L. Sabran, G. E. Mark, R. V. Guntaka, and J. M. Taylor. 1978. Analysis of unintegrated avian RNA tumor virus double-stranded DNA intermediates. *J. Virol.* **28**:810-818.
  14. Lavelle, G., L. Foote, R. L. Heberling, and S. S. Kalter. 1979. Expression of baboon endogenous virus in exogenously infected baboon cells. *J. Virol.* **30**:390-393.
  15. McAllister, R. M., W. A. Nelson-Rees, E. Y. Johnson, R. W. Rongey, and M. B. Gardner. 1971. Disseminated rhabdomyosarcomas formed in kittens by cultured human rhabdomyosarcoma cells. *J. Natl. Cancer Inst.* **47**:603-611.
  16. McAllister, R. M., M. O. Nicolson, R. Heberling, H. Charman, N. Rice, and R. V. Gilden. 1978. Infectivity of endogenous baboon type-C virus-related genes, p. 135-138. *In* P. Bentvelzen, J. Hilgers, and D. S. Yohn (ed.), *Advances in comparative leukemia research*. Elsevier/North-Holland Biomedical Press, Amsterdam, New York.
  17. Rothenberg, E., and D. Baltimore. 1976. Ordered transcription of RNA tumor virus genomes. *J. Mol. Biol.* **106**:109-131.
  18. Shank, P. R., J. C. Cohen, H. E. Varmus, K. R. Yamamoto, and G. M. Ringold. 1978. Mapping of linear and circular forms of mouse mammary tumor virus DNA with restriction endonucleases: evidence for a large specific deletion occurring at a high frequency during circularization. *Proc. Natl. Acad. Sci. U.S.A.* **75**:2112-2116.
  19. Shank, P. R., S. H. Hughes, H.-J. Kung, J. E. Majors, N. Quintrell, R. V. Guntaka, J. M. Bishop, and H. E. Varmus. 1978. Mapping unintegrated avian sarcoma virus DNA: termini of linear DNA bear 300 nucleotides present once or twice in two species of circular DNA. *Cell* **15**:1383-1395.
  20. Shank, P. R., and H. E. Varmus. 1978. Virus-specific DNA in the cytoplasm of avian sarcoma virus-infected cells is a precursor to covalently closed circular viral DNA in the nucleus. *J. Virol.* **5**:104-114.
  21. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**:503-517.
  22. Sutcliffe, J. G. 1978. pBR322 restriction map derived from the DNA sequence: accurate DNA size markers up to 4361 nucleotide pairs. *Nucleic Acids Res.* **5**:2721-2728.
  23. Taylor, J. M., and R. Illmensee. 1975. Site on the RNA of an avian sarcoma virus at which primer is bound. *J. Virol.* **16**:553-558.
  24. Taylor, J. M., R. Illmensee, and J. Summers. 1976. Efficient transcription of RNA into DNA by avian sarcoma virus polymerase. *Biochim. Biophys. Acta* **442**:324-330.
  25. Todaro, G. J., C. J. Sherr, R. E. Benveniste, M. M. Lieber, and J. L. Melnick. 1974. Type C viruses of baboons: isolation from normal cell cultures. *Cell* **2**:55-61.
  26. Varmus, H. E., S. Heasley, H.-J. Kung, H. Oppermann, V. C. Smith, J. M. Bishop, and P. R. Shank. 1978. Kinetics of synthesis, structure and purification of avian sarcoma virus-specific DNA made in the cytoplasm of acutely infected cells. *J. Mol. Biol.* **120**:55-82.
  27. Yoshimura, F. K., and R. A. Weinberg. 1979. Restriction endonuclease cleavage of linear and closed circular murine leukemia viral DNAs: discovery of a smaller circular form. *Cell* **16**:323-332.