Conservation of Gene Organization in the Lymphotropic Herpesviruses Herpesvirus Saimiri and Epstein-Barr Virus

U. A. GOMPELS,[†] M. A. CRAXTON, AND R. W. HONESS*

Division of Virology, National Institute for Medical Research, London NW7 1AA, United Kingdom

Received 17 September 1987/Accepted 5 November 1987

By analyses of short DNA sequences, we have deduced the overall arrangement of genes in the (A+T)-rich coding sequences of herpesvirus saimiri (HVS) relative to the arrangements of homologous genes in the (G+C)-rich coding sequences of the Epstein-Barr virus (EBV) genome and the (A+T)-rich sequences of the varicella-zoster virus (VZV) genome. Fragments of HVS DNA from 13 separate sites within the 111 kilobase pairs of the light DNA coding sequences of the genome were subcloned into M13 vectors, and sequences of up to 350 bases were determined from each of these sites. Amino acid sequences predicted for fragments of open reading frames defined by these sequences were compared with a library of the protein sequences of major open reading frames predicted from the complete DNA sequences of VZV and EBV. Of the 13 short amino acid sequences obtained from HVS, only 3 were recognizably homologous to proteins encoded by VZV, but all 13 HVS sequences were unambiguously homologous to gene products encoded by EBV. The HVS reading frames identified by this method included homologs of the major capsid polypeptides, glycoprotein H, the major nonstructural DNA-binding protein, thymidine kinase, and the homolog of the regulatory gene product of the BMLF1 reading frame of EBV. Locally as well as globally, the order and relative orientation of these genes resembled that of their homologs on the EBV genome. Despite the major differences in their nucleotide compositions and in the nature and arrangements of reiterated DNA sequences, the genomes of the lymphotropic herpesviruses HVS and EBV encode closely related proteins, and they share a common organization of these coding sequences which differs from that of the neurotropic herpesviruses, VZV and herpes simplex virus.

A subdivision of the herpesvirus family into three major subgroupings, the alpha-, beta-, and gammaherpesviruses, has been suggested on the basis of differences in biological properties (31, 33, 53). The alphaherpesviruses are neurotropic viruses typified by herpes simplex virus (HSV) and varicella-zoster virus (VZV), the betaherpesviruses are the cytomegaloviruses or salivary gland inclusion agents, and the gammaherpesviruses are lymphotropic viruses exemplified by the B-lymphotropic Epstein-Barr virus (EBV) and the T-lymphotropic herpesvirus saimiri (HVS). Our studies aim to define the relationship among the widely divergent biological properties of herpesviruses and the differences in their gene products and their expression. In particular, our objectives are to understand the distinctions between the neurotropic and lymphotropic viruses. We therefore chose to study the molecular biology of HVS. This virus has some relatively clearly defined biological properties combined with a distinctive genome structure. Thus, the virus can be isolated from T lymphocytes of healthy adult members of its natural host species (*Saimiri sciureus*) but causes fatal T-cell lymphoproliferative disease in other New World monkeys (i.e., marmosets and owl monkeys) (25). The virus is also capable of immortalizing a subset of the T lymphocytes from marmoset peripheral blood after infection in vitro (22). The virus genome in tumor tissues and in the cell lines established by in vitro immortalization persists as an extensively methylated, multicopy episome with restricted virus gene expression (21, 25). These latter characteristics are clearly

reminiscent of EBV in its association with the B lymphocyte (34). However, the genomes of HVS and EBV differ markedly in nucleotide composition and in the nature and arrangement of repetitive DNA sequences. The HVS genome is about 160 kilobase pairs (kbp), 111 kbp of which are coding sequences of 36% G+C bounded by terminal repetition of a 1.44-kbp direct repeat of 71% G+C (3, 25, 63). In contrast, the EBV genome is about 172 kbp, most of which is a coding sequence with an average G+C composition of 60%, bounded by terminal direct repeats of 0.5 kbp. In addition, the EBV genome contains many sets of internal direct repeats, the largest being the *Bam*HI-W repeats of more than 3 kbp reiterated up to 12 times (1).

Despite these differences in overall genome composition and organization, serological analyses and comparisons between virion polypeptides and nonstructural proteins (DNA-binding proteins) have shown that the proteins encoded by the (G+C)-rich EBV and the (A+T)-rich HVS lymphotropic gammaherpesviruses reveal more similarities than those seen in comparisons with the neurotropic alphaherpesviruses such as HSV and VZV (7, 45, 49).

The complete sequences of the genomes of EBV (1) and VZV (14) and large samples of sequences from HSV (13, 15, 42, 48) and human cytomegalovirus (36, 66) have recently been reported. Analyses of the major open reading frames predicted from these sequences have permitted the recognition of genes encoding many proteins which are common to all these viruses as well as the recognition of genes specific to particular viruses or subgroups of viruses. More significantly, the order and orientation of homologous genes is closely similar throughout the genomes of VZV and HSV (13, 17), but these global arrangements of homologous genes

* Corresponding author.

[†] Present address: Division of Virology, Department of Pathology, University of Cambridge, Cambridge, United Kingdom.

are not preserved in the genomes of EBV (1, 16) or human cytomegalovirus (36). In previous studies, we showed that the rightmost gene of HVS, (the 160-kilodalton (kDa) tegument protein, was homologous to the leftmost gene of EBV, the 140-kDa product of the BNRF1 reading frame (9), and that no obvious homologs of these genes were present in VZV. However, we have also shown that a gene with no homolog in the EBV genome (the thymidylate synthase gene) but with a very high degree of homology to other thymidylate synthase genes, including one in VZV (64), is also located to the right of the HVS genome (32; see Fig. 1).

In this report, we present the results of an analysis of 13 short nucleotide sequences from subclones distributed across the (A+T)-rich light DNA (L-DNA) sequences of HVS, which have permitted the rapid establishment of important aspects of the genetic organization of HVS. In particular, we show that HVS has an arrangement of its protein coding sequences which is grossly colinear with that of the (G+C)-rich gammaherpesvirus EBV.

MATERIALS AND METHODS

Enzymes and nucleotides. Restriction enzymes, T4 DNA ligase, DNA polymerase (Klenow fragment), deoxynucleotides, and dideoxynucleotides were purchased from Boehringer Corp. (London) Ltd.

Recombinant plasmids. Cloned *EcoRI* I, *EcoRI* E, *EcoRI* F, *EcoRI* G, *EcoRI* B, and *KpnI* F fragments of the HVS strain 11 genome (23) used were as described by Knust et al. (35). The *EcoRI* fragments were cloned into the plasmid vector pACYC184, whereas the *KpnI* F fragment was cloned into the cosmid vector pJC81. The recombinant plasmid p*EcoRI*-(I/E') consists of *EcoRI*-I plus the adjacent 1.35-kbp *EcoRI*-to-*HindIII* fragment of *EcoRI*-E cloned into pUC8 and was provided by J. Nicholas, National Institute for Medical Research, London, United Kingdom.

M13 subcloning and DNA sequencing. Nucleotide sequences were determined by the dideoxynucleotide chain termination method (54, 55), essentially as described by Bankier and Barrell (2), with sequences cloned into bacteriophages M13mp18 and M13mp19 (44, 68). The recombinant plasmids or cosmid were digested with *Bam*HI, *EcoRI*, *Pst*I, or *HindIII*, and appropriate fragments were purified by agarose gel electrophoresis, followed by extraction by the glass filter method (10, 65). Restriction fragments were cloned into the homologous sites in the polylinkers of M13mp18 or M13mp19, and their origins and orientations were confirmed by *Msp*I, *Kpn*I, *Pst*I, *EcoRI*, or *HindIII* digestion of the double-stranded replicative-form DNA. Small-scale preparations of replicative-form DNA were purified by the method of Birnboim and Doly (6) from 1.5-ml cultures of bacteria infected with the recombinant phages as described for the preparation of single-stranded DNA preparations (2).

Dideoxy sequencing from single-stranded templates was performed with a 17-mer universal primer and with [³⁵S]thio-dATP (Amersham International, Amersham, United Kingdom) as a radioactive label, and the products were fractionated on Tris-borate-EDTA buffer-gradient polyacrylamide gels (4). The data presented in this report are of nucleotide sequences of up to 350 bases in length, determined for each M13 subclone as the result of a single fractionation of the products from the polymerization reaction. Since the HVS genome is (A+T)-rich, ambiguities such as those introduced by compressions in (G+C)-rich sequences were not often

observed, but because the present sequences were derived from only one strand of the DNA, some errors may be present. Since our objective was to provide a rapid method to produce a genetic map, a high level of reliability in the sequences obtained was sufficient to achieve this objective.

Computer analyses of the sequence data. Sequence data were recorded and analyzed on a DEC 20-60 computer (Digital Equipment) by using the programs of Staden (57-62). Open reading frames were identified and translated with the ANALYSEQ program package (61, 62). Reading frames greater than 150 nucleotides in length from the HVS sequences or their complements were examined and selected for coding potential by the positional base preference and codon usage methods (59, 60, 62). Amino acid sequences predicted from these open reading frames were compared with a library of translated open reading frames derived from the nucleotide sequence of the EBV and the VZV genomes (1, 14) by use of the Wilbur and Lipman (67) protein similarity search strategy. The program used was implemented by R. Mott in the Molecular Genetics and Sequencing program package, developed at the National Institute for Medical Research (unpublished results). This method uses a score matrix based on amino acid substitutions in families of related proteins, MDM78 (18), modified by Staden (57), who added 10 to each score so that all scores were greater than zero. Scores were calculated from the sums of identical or similar matches, with a match set derived from McLachlan (43), minus a penalty for insertion of gaps necessary for maximal alignments. The highest scoring alignments with an overall similarity of greater than 40% were analyzed further by dot matrix comparisons similar to that of DIAGON (57), as implemented in the Molecular Genetics and Sequencing program package by P. Gillett at the National Institute for Medical Research.

Alignments of the HVS amino acid sequences to sequences in the EBV, VZV, or random sequence (see below) libraries were made by using a *K*-tuple of 2 and a stringent gap penalty of 90, where a score of 10 indicates a match representing an amino acid substitution neither favored or disfavored. The highest-scoring alignments between the HVS and EBV amino acid sequences selected by this method were then optimally aligned by using a *K*-tuple of 1 and a gap penalty of 30. In the first selection from the EBV library, the highest-scoring alignments had greater than 40% similarity with no insertions. In the final alignments of these 13 homologous amino acid sequences, seven required no insertions, and there were no more than three insertions in each of the remaining six alignments.

The random amino acid sequences used for significance testing were generated by the PROTRN program, implemented by R. Mott at the National Institute for Medical Research. Random sequences of the same size as each tested HVS sequence were created by using amino acid frequencies deduced from the codon usage for EBV, as compiled by Maruyama et al. (39) from open reading frames identified by Baer et al. (1). For a number of amino acids, these frequencies differed significantly from those deduced from codon usage tables compiled by Maruyama et al. (39) for 135 human genes within GenBank; these differences could be accounted for by the higher relative G+C content of EBV DNA. The major differences included increased frequencies of proline, alanine, and glycine and decreased frequencies of lysine and asparagine. Thus, as an additional guide for significance of a match to an EBV amino acid sequence, the weight matrix score for an identical amino acid residue match can be divided by its frequency and the following order of signifi-

cance (high to low) can be obtained: W, C, M, Y, H, F, K, I, N, Q, R, D, E, V, P, T, G, S, L, and A.

RESULTS

Derivation of clones and nucleotide sequence determination. Fragments of DNA distributed throughout the HVS genome (Fig. 1) were cloned directly into M13 vectors, and nucleotide sequences of up to 350 bases were determined from each cloned fragment. These subclones were derived from either ligation of vectors with mixtures of DNA fragments from digests of cloned HVS DNA and then mapping the resulting random clones by known restriction enzyme recognition sites or from ligation of purified HVS DNA fragments of known map location with vector DNA. Thus, the sequence from *KpnI* F was derived from subclones 1 and 2, which were produced by digestion of the *KpnI* F cosmid with *PstI* and ligation of the resulting mixture of fragments with *PstI*-digested M13mp18. Similarly, subclones 6, 7, 8, and 9 were produced from the digestion of *PstI*-digested *EcoRI*-B plasmid with *PstI*-digested M13mp18. The *BamHI* F and *EcoRI* I DNA fragments were purified and ligated directly with either *BamHI*-digested or *EcoRI*-digested M13mp18 to produce clone 10 (*BamHI*-F) and clones 11 and 12 (either orientation of the *EcoRI* I inserted DNA). Additional clones 3, 4, 5, and 13 were made by directed ligations into double-digested M13mp18 or M13mp19. Accordingly, clone 3 was generated by ligation of a *PstI*-to-*HindIII* subfragment of *EcoRI*-F with *PstI*- and *HindIII*-digested M13mp18 vector; clones 4 and 5 were generated by ligation of the 2.7-kbp *PstI*-to-*EcoRI* subfragment of the *EcoRI* B plasmid with either *PstI*- and *EcoRI*-digested M13mp19 (clone 4) or M13mp18 (clone 5); and clone 13 was generated by ligation of the 1.35-kbp *EcoRI*-to-*HindIII* subfragment from plasmid pEcoRI-(I/E') with *EcoRI*- and *HindIII*-digested M13mp19.

Sequence analyses. Open reading frames were identified for nucleotide sequences derived from each of the HVS M13 recombinant clones. These had a higher-than-average probability of being coding sequences as assessed by the posi-

tional base preference method or by codon usage analyses with a reference codon usage table compiled from HVS thymidylate synthase and 160-kDa protein genes (9, 32). The reading frames were designated HVS RF1 through RF13, according to the respective HVS M13 clones from which the nucleotide sequences were derived.

Amino acid sequences encoded by HVS RF1 through RF13 were compared with libraries of translated VZV and EBV open reading frames. Only 3 of the 13 HVS sequences were clearly homologous with VZV gene products (HVS RF3 to VZV RF34, HVS RF8 to VZV RF40, and HVS RF13 to VZV RF4, with overall similarities of 42%, 46%, and 42%, respectively), whereas all 13 HVS reading frames encoded homologous products to EBV sequences with greater than 40% similarity. The alignments between the HVS and EBV amino acid sequences are shown (Fig. 2). Also shown is the gene fragment adjacent to the thymidylate synthase gene (8, 32; K. R. Cameron and R. W. Honess, unpublished results), labeled HVS RF14, which encodes sequences homologous to those encoded by EBV BFLF2. The overall similarities of the 14 alignments shown ranged between 45 and 72%.

Significance of homologous amino acid sequences. All the amino acid sequence alignments shown in Fig. 2 have similarities clearly higher than the similarities produced by chance among unrelated proteins (<30%) (18). However, the amino acid sequences analyzed (50 to 116 residues) are much smaller than the average protein, and the alignments obtained are to sequences with a biased amino acid composition influenced by the effects of the higher-than-average G+C content of the EBV genome on codon usage. The significance of the observed similarities was therefore assessed by two additional methods. First, the score obtained for the best alignment of the HVS sequences with the EBV sequences was compared with the 10 next-best scores selected from a search of the library composed of amino acid sequences from the open reading frames of EBV. Second, the scores of the proposed best alignments were compared with scores obtained for EBV-like sequences generated by random choice of residues by using the amino acid frequencies observed in EBV. In each case, the scores of the

FIG. 1. Locations of M13 subclones (numbered 1 to 13) relative to selected endonuclease cleavage sites and previously sequenced genes (thymidylate synthase and the 160K-gene; 9, 32) in the L-DNA component of HVS strain 11 Onc DNA. The upper line shows the 111-kbp L-DNA component of the HVS genome in its conventional orientation, with the 160K-gene (in *KpnI* + *SmaI* E) at the right and the $>30 \times 1.4$ -kbp reiterations composing the terminal heavy DNA component arbitrarily drawn at the right. Locations of the *KpnI* F, *EcoRI* F, *EcoRI* B, *BamHI* F, and *HindIII* A (*EcoRI* I) fragments are shown, with sources of subclones and sites from which sequence data were obtained (●→, direction of sequencing) presented on expanded maps of these fragments (P, *PstI* sites; for other enzymes, only those sites relevant to the cloning strategy or the orientation of cloned fragments are shown). Note that, with the exception of clone 10, the orientation of the HVS DNA fragments within the vector sequence was determined with respect to mapped restriction enzyme recognition sites.

BALF2

GAGCTTCAACATCTTTTGACAACCTAAAA

E	L	Q	H	L	L	H	N	L	K
E	L	V	H	L	L	E	N	S	R

180

BALF 3

..... CCTATTGCA
- - P I A
A S P V A 568

BVRF1

AGAGGC

R	G
R	G

193

BVRF1

AGTGCTATCAAACCA
S A I K P
P A I R S 586

BXLF1

TACAGACCTGAAACAAATGTACATATTCCT

Y	R	P	E	T	N	V	H	I	P
Y	K	T	S	G	T	L	G	I	A

85

BXLF2

CACACAGTAAGAAGA
H T V R R
S C I N R 504

BcRF1

AAGTGTAGTAAATGTCATAGGC ... CAGGCTGATAACTCTATTATG

K	L	V	N	V	I	G	-	Q	A	D	N	S	I	M
Q	L	L	T	F	L	G	E	E	A	D	N	Q	I	M

303

BcLF1

FIG. 2. Identification of open reading frames from the nucleotide sequences of HVS M13 recombinant clones 1 through 13. RF4, RF5, RF9, and RF12 were from the complement of the nucleotide sequence from clones 4, 5, 9, and 12. Beneath each nucleotide sequence (top line of each panel), the predicted HVS amino acid sequences are shown (middle lines); these are aligned with amino acid sequences encoded by portions of the EBV open reading frames indicated (bottom lines; EBV reading frames are identified below each alignment by the nomenclature of Baer et al. [1]). Gaps are introduced in some sequences to maximize the alignments, and identical and similar (i.e., I≡L≡V≡M, D≡E; R≡K) amino acid sequences are boxed. RF14 is from nucleotide sequence analyses adjacent to the thymidylate synthase gene (Fig. 1) (8, 32; Honess and Cameron, unpublished).

TABLE 1. Comparison of HVS reading frames and EBV library sequences

HVS sequence	Best EBV matches			10 Next-best EBV matches		10 Random EBV matches	
	EBV open reading frame ^a	Score	% Similarity	Mean score $\pm$ SD	% Similarity	Mean score $\pm$ SD	% Similarity
RF1	BALF2	591	76	147 $\pm$ 10	24	74 $\pm$ 20	12
RF2	BALF3	435	59	192 $\pm$ 11	23	111 $\pm$ 22	16
RF3	BVRF1	636	58	203 $\pm$ 24	23	99 $\pm$ 20	13
RF4	BVRF1	542	51	187 $\pm$ 14	23	103 $\pm$ 28	13
RF5	BXLF1	317	48	169 $\pm$ 10	24	94 $\pm$ 27	15
RF6	BXLF2	275	50	163 $\pm$ 8	27	84 $\pm$ 19	15
RF7	BcRF1	836	74	208 $\pm$ 6	20	113 $\pm$ 33	16
RF8	BcLF1	626	64	224 $\pm$ 14	26	130 $\pm$ 33	16
RF9	BcLF1	616	53	222 $\pm$ 14	22	139 $\pm$ 19	13
RF10	BDRF1	768	61	226 $\pm$ 9	20	125 $\pm$ 19	13
RF11	BSLF1	633	51	198 $\pm$ 23	25	114 $\pm$ 21	14
RF12	BMLF1	326	40	207 $\pm$ 12	25	121 $\pm$ 21	15
RF13	BMLF1	271	51	146 $\pm$ 13	25	80 $\pm$ 28	14
RF14	BFLF2	390	42	223 $\pm$ 10	20	134 $\pm$ 26	16

^a EBV reading frames are identified by the nomenclature applied by Baer et al. (1).

proposed best alignments were significantly higher than either of the test scores (Table 1).

The alignments between HVS and EBV amino acid sequences (Fig. 2) were ranked by overall percent similarity and the number of standard deviations above the mean score of alignments of the HVS sequences with the 10 next-best scores from the EBV library or with 10 randomly generated EBV-like sequences (Table 2). The numbers of standard deviations above the mean range from 7 to 105, and the percent similarities range from 46 to 72%. These results are significant by any reasonable criteria (see discussions in references 18, 24, and 27). Thus, even the lowest-scoring alignments between EBV and HVS amino acid sequences (HVS RF12, 13, and RF6 with EBV BMLF1 and BXLF2, Table 2) permitted the location and subsequent analyses of the HVS homologs of these important regulatory protein and glycoprotein genes (U. A. Gompels, M. A. Craxton, and R. W. Honess, manuscript in preparation).

HVS homologs of proteins of other herpesviruses with known function. Many of the EBV genes homologous with HVS RF1 to RF14 encode products which have been shown to be similar to products of both VZV and HSV (Table 3).

Thus, the products of EBV genes BMLF1 and BXLF2, homologous with HVS RF12, RF13, and RF6, as described above, are conserved in VZV as well as in HSV, in which they are characterized as the HSV immediate-early 63,000-molecular-weight protein (IE 63K protein) and glycoprotein H, respectively (14, 16, 29, 41). Similarly, EBV BALF2, BXLF1, and BcLF1, encoding homologs of HVS RF1, RF5, and RF8 plus RF9, respectively, are also conserved in VZV and HSV. In HSV they are further characterized as the major DNA-binding protein, thymidine kinase, and major capsid protein (15, 16, 38, 48). Thus, the locations of the HVS homologs to these genes have been identified.

Although direct comparisons of the HVS reading frame products to a library of translated VZV open reading frames (14) showed only three clearly significant alignments (Table 3), comparisons of the homologous short EBV amino acid sequences with the same library showed only the same three alignments plus an alignment of 45% similarity to the major DNA-binding protein homolog of VZV. With the exception of BcRF1, all the EBV open reading frames listed in Table 3 have been shown to encode proteins homologous to the listed VZV gene products (16); however, the regions encod-

TABLE 2. Ranking of observed similarities between HVS RF1 to RF14 and their EBV homologs

HVS RF	No. of amino acids	EBV homolog	Optimal alignment in EBV homolog	% Similarity	EBV alignment significance ^a	Random alignment significance ^b
RF7	85	BcRF1	420-504	72	105	56
RF1	53	BALF2	442-494	72	59	26
RF10	108	BGRF1	135-206	54	60	34
RF8	92	BcLF1	209-303	69	29	15
RF3	83	BVRF1	484-568	62	18	27
RF9	105	BcLF1	742-873	53	28	18
RF4	81	BVRF1	112-193	56	25	16
RF2	70	BALF3	111-180	58	22	15
RF11	100	BSLF1	638-737	46	19	25
RF14	116	BFLF2	192-307	52	17	10
RF5	64	BXLF1	522-586	52	15	8
RF6	50	BXLF2	36-85	46	14	10
RF13	51	BMLF1	406-455	51	10	7
RF12	80	BMLF1	311-388	46	7	10

^a Number of standard deviations by which the score for the optimal alignment exceeds the score for the ten next-best alignments with EBV sequences.

^b Number of standard deviations by which the score for the optimal alignment exceeds the score for alignment with 10 random sequences which have the same amino acid frequencies as the EBV homolog.

TABLE 3. Functions associated with homologs of HVS RF1 to RF14 from studies with EBV or with VZV and HSV

HVS RF ^a	EBV homolog	VZV homolog (16)	HSV function
RF1	BALF2	RF29	Major DNA-binding protein ICP 18.5K protein (47)
RF2	BALF3	RF30	
RF3 plus RF4	BVRF1	RF34	
RF5	BXLF1	RF36	Thymidine kinase (38) Glycoprotein H (29, 41)
RF6	BXLF2	RF37	
RF7	BcRF1		
RF8 plus RF9	BcLF1	RF40	Major capsid protein (15) Spliced gene product (12)
RF10	BGRF1	RF45	
RF11	BSLF1	RF6	
RF12 plus RF13	BMLF1	RF4	IE 63K protein (16)
RF14	BFLF2	RF27	

^a RF3, RF8, and RF13 showed >40% similarity in direct comparison with the reading frames of VZV.

ing amino acid sequences homologous to those of the HVS reading frames show similarities to the VZV sequences of only 30 to 47%. A similar range was observed for comparisons of the HVS reading frame-encoded sequences directly to the proposed homologous sequences in VZV. In contrast, the respective EBV and HVS sequences show similarities of 46 to 72% (Table 2).

Arrangement of the HVS reading frames. In Table 3, the HVS reading frames are listed in the order in which they appear in the genome (Fig. 1). The corresponding homologous EBV open reading frames appear in the same order on the EBV genome (1). Thus, the relative arrangement of these HVS genes is similar to that of their EBV counterparts. In contrast, the VZV genes, also numbered to reflect their order on the genome (14), are rearranged with respect to the HVS reading frames (Table 3).

A more detailed analysis of the spacing and orientation of the HVS reading frames with respect to homologous regions of the EBV genome is shown in Fig. 3 and corresponds very closely with that of the homologous genes of EBV. In this figure, the position and orientation of the HVS reading frame on the genome is shown in relation to the map coordinates of the parts of the EBV open reading frames encoding homologous sequences. With the exception of the HVS homolog to EBV BGRF1, the orientations of all the HVS reading frames

have been determined. Initially, this was deduced by identification of the direction of the inserted HVS DNA in the M13 recombinant clones from which the nucleotide sequences of the respective HVS reading frames were derived (Fig. 3). Thus, HVS RF3, RF4, RF5, and RF13 were derived from M13 recombinants generated by directed cloning into M13mp18 and M13mp19 vectors, whereas HVS RF2, RF8, RF9, RF11, and RF12 were oriented by identification of mapped restriction sites (63) in the respective nucleotide sequences (Fig. 1 and 3). In the case of HVS RF2 derived from clone 2, a *Pst*I fragment composed of HVS DNA and vector sequences, restriction enzyme sites within the vector sequences were used to determine the orientation of the inserted DNA. The arrangement of these HVS reading frames matched the arrangement of the homologous regions of the EBV genome encoding the products of BVRF1, BXLF1 (thymidine kinase), BcLF1 (major capsid protein), BSLF1, and BMLF1 (homolog of the IE 63K protein) (Fig. 3). The order of the other HVS reading frames was consistent with homologous regions of the EBV genome; the actual orientations could not be shown as above (Fig. 3). However, the orientations of HVS RF1, RF6, and RF7 (Fig. 3), encoding fragments of the HVS major DNA-binding protein and glycoprotein H, were confirmed by further nucleotide sequence analyses. The complete nucleotide sequence has been determined for the HVS homologs of EBV BXLF1, BXLF2, and BMLF1 (Gompels et al., in preparation), confirming the orientation of HVS RF6 and RF7, as well as RF5, RF12, and RF13 (Fig. 3). Additional sequence analysis has also shown that the *Pst*I to *Kpn*I 3.0-kbp fragment at the left of *Pst*I-R (Fig. 1, clone 1a) encodes the N-terminal portion of the HVS gene product homologous with that of EBV BALF2 (J. Nicholas, unpublished results). In summary, it is apparent that the arrangement of HVS gene fragments 1 to 13 (Fig. 3) corresponds very closely with that of the homologous genes of EBV.

Conservation of HVS arrangement compared with that of EBV and VZV. The HVS reading frames are homologous to EBV open reading frames, which can be grouped into three blocks on the basis of comparisons between EBV and the alphaherpesvirus VZV. These blocks of genes are regions of the genomes within which there is a conserved order of genes, many encoding conserved gene products (summarized in Fig. 3 and 4 as gene blocks 1 through 4). The spacing between some of the gene blocks in EBV and HVS may

FIG. 3. Relationships among the spacing and orientation of HVS RF1 to RF13 in the L-DNA component of HVS and their homologs on the genome of the B95-8 strain of EBV. (A) Maps of the relevant restriction endonuclease cleavage sites (K, *Kpn*I; P, *Pst*I; e, *Eco*RI; h, *Hind*III; m, *Msp*I; s, *Sal*I; b, *Bam*HI; pv, *Pvu*II) in the L-DNA component of HVS. (B) Positions of HVS RF1 to RF13 relative to these maps, with the orientation of the reading frames indicated by arrowheads (N terminus to C terminus). (C) Organization of the complete open reading frames of the EBV genome (in the inverse of the prototype orientation) which contain homologs of HVS RF1 to RF13, aligned with these homologs. Note that the same linear scales (centimeter/kilobase pair) are on panels A, B, and C and that over these regions of the two virus genomes the order, orientation, and spacing of homologous reading frames are congruent. (D) Symbolic coding of these three blocks of homologous genes as used in Fig. 4; symbols: , block 3; , block 4; , block 1.

FIG. 4. Comparison of the order, orientation, and spacing of groups or blocks of homologous genes on the mature, linear genomes of VZV (an alphaherpesvirus), EBV, and HVS. The open reading frames (small solid arrows, arrowheads to C termini) of VZV and EBV may be conveniently considered as members of four major groups or blocks of genes (blocks 1, 2, 3, and DNA polymerase and block 4, indicated by the large open or shaded arrows) within which clearly homologous gene products are in the same order and have the same relative orientation and similar spacing. Only those regions bounded by open reading frames encoding homologous amino acid sequences (16) are shown. Open reading frames in VZV or EBV within a conserved block of genes but without detectable homologs are marked by a dot. Thus block 1 (▨) contains RF4 to RF8 of VZV and includes homologs of the IE 63K protein and dUTPase (dU) of HSV type 1. Block 2 (▩) contains RF18 to RF27 of VZV and includes homologs of the ribonucleotide reductase (RR) polypeptides of HSV type 1 (26). Block 3 (□) contains RF29 to RF31 of VZV and includes homologs of the HSV type 1 major DNA-binding protein (DBP) and glycoprotein B (gB; 46) with DNA polymerase (POL; VZV RF28) at the 5' end of block 3 in VZV and the 3' end in EBV. Block 4 (▤) contains RF18 to RF27 of VZV and thus includes homologs of the thymidine kinase (TK), glycoprotein H (gH), major capsid polypeptide (MCP), and alkaline exonuclease (XO) genes of HSV type 1 (16). Also shown are the positions of the thymidylate synthase genes of VZV and HVS (TS) and the gene for the 140K nonglycosylated membrane antigen of EBV and the 160-kDa tegument protein which is its homolog in HVS (MA). Note that HVS RF14 encodes the homolog of EBV BFLF2 and thus VZV RF27 (Fig. 2, Table 3), which are the terminal members of gene block 2. Shown below each summary of the organization of these gene blocks in the three viruses is a schematic summary of some of the major structural features of their genomes. Tandem direct repeats are indicated by vertical lines or rectangles; TR, terminal direct repeats of EBV and HVS; IRL, IRS, and TRS, terminal and internal inverted repeats of VZV; O, origins of DNA replication (1, 14); DL and DR, internal duplications of the EBV genome. As in Fig. 3, the orientation of the EBV genome in this diagram is inverted relative to the conventional orientation.

differ. In particular, the distance between the nucleotide sequence derived from the HVS clones 4 and 13 on the HVS genome is 44.7 kbp, whereas the corresponding sites on the EBV genome (from the end of BMLF1 in gene block 1 to the beginning of BVRF1 in gene block 4) are separated by 62.7 kbp. Thus, there is an 18-kbp discrepancy between the spacing of the predicted HVS gene blocks 4 and 1 versus their counterparts in EBV. Approximately 7 kbp can be accounted for by the presence of repetitive DNA and gene duplications in the EBV genome (Fig. 4), leaving a 10-kbp discrepancy. The overall complexity of unique DNA sequences in EBV strain B95-8 is approximately 124 kbp compared with an estimated 110 kbp for HVS. Thus, the major difference in the unique sequence content of these two genomes could be accounted for by a deletion of sequences between gene blocks 4 and 1 of HVS relative to the sequences of EBV.

Although there are some differences in spacing between sets of conserved genes in EBV and HVS, their general

organization remains similar. In comparisons of the overall arrangements of these conserved groups of genes between gammaherpesviruses HVS and EBV and alphaherpesviruses VZV and HSV, groups that are located near termini in one class are located internally in the other (Fig. 4). HVS RF1 and RF2 are homologous to genes in EBV which are part of a group represented by the conserved major DNA-binding protein gene located in EBV and HVS toward a terminus, whereas in VZV and HSV it is located internally (Fig. 4). In contrast, the adjacent block of conserved genes in HVS (represented by fragments encoding thymidine kinase, glycoprotein H, and major capsid protein homologs) is located internally in both EBV and HVS genomes, whereas in VZV and HSV it extends toward a terminus of U_L . Moreover, the relative orientation of these genes in the gammaherpesviruses is the reverse of that seen in the alphaherpesviruses. The rearrangements of the EBV gene order relative to those of HSV and VZV have been documented (16, 41, 42, 48). The arrangement of the HVS gene fragments discussed here

indicates an overall genetic organization that is grossly colinear with the gammaherpesvirus EBV and distinct from that of the alphaherpesviruses VZV and HSV (Fig. 4).

DISCUSSION

This report describes the overall genetic colinearity between the genomes of the lymphotropic gammaherpesviruses HVS and EBV. Small nucleotide sequences of up to 350 nucleotides were derived from 13 clones of HVS DNA with known map locations. Open reading frames within these HVS nucleotide sequences or their complements encoded short amino acid sequences which were homologous to portions of gene products encoded by the EBV genome. All the homologies were significant and were used not only to demonstrate the overall conservation of gene order between EBV and HVS but also to locate HVS genes encoding homologs of proteins from other herpesviruses whose functions are known, i.e., the HSV IE 63K protein, glycoprotein H, the major capsid polypeptide, thymidine kinase, and the major nonstructural DNA-binding protein (Table 3, Fig. 4).

Methods used to assess the significance of similarities between two sequences have involved either comparisons with a sufficiently large data base or comparisons between randomly reassorted versions of the test sequences (27). We have used both approaches here, except that EBV amino acid frequencies were used in generating artificial test sequences such that the influence of compositional bias on observed similarities was assessed. This was successful in showing significant similarities between amino acid sequences of only 50 to 116 residues in length. Accordingly, the alignments shown here of the partial gene products conserved between EBV and HVS range from low to high (Table 2) as do the alignments of complete gene products conserved between EBV and VZV (16). For example, EBV BMLF1 and VZV RF4 are among the least similar of the homologs detected in comparisons of EBV sequences with VZV sequences. HVS RF12 with BMLF1 is the lowest-ranked similarity in comparisons between HVS and EBV and shows no clearly significant similarity with VZV RF4. However, subsequent sequencing of the entire HVS gene identified by RF12 has shown a gene product which is clearly homologous to those of both EBV BMLF1 and VZV RF4 (unpublished results). The amino acid sequence encoded by HVS RF8 has a relatively higher degree of similarity to the respective portion of the product of EBV BcLF1 (Tables 2 and 3) and was also homologous to sequences encoded by VZV RF40. In contrast, the most similar sequences found in comparisons between the HVS and EBV proteins were between products of HVS RF7 and EBV BcRF1, and these proteins have no detectable counterpart encoded in the VZV genome (Table 3, Fig. 3 and 4; 16). These observations suggest two points. First, a minimum overall similarity in the amino acid sequences of the test and query sequences is necessary before comparisons of short amino acid sequences such as those described here could be successful in determining conserved gene order or location. The similarities observed in comparisons between the gammaherpesviruses HVS and EBV were greater than 40%, whereas in comparisons of these protein fragments from HVS or from EBV with those of the alphaherpesvirus VZV, only three had similarities of more than 40%. Thus, although HVS has an (A+T)-rich genome (36% G+C), the HVS reading frames analyzed here encoded gene products which were more similar to those of the (G+C)-rich EBV (60% G+C) from the same subgroup than to those of the (A+T)-rich VZV genome

(46% G+C) from the alphaherpesvirus subgroup. Second, although some gene products are subgroup specific, others, while conserved among herpesviruses, have different domains which may be subgroup specific.

Previous studies of proteins specified by HVS are consistent with the present identification of HVS genes encoding products homologous to those of EBV. For example, Hell et al. (30) applied hybrid-selected in vitro translations to map some 45 virus-induced polypeptides, ranging in size from 12 to 152 kDa, to the HVS genome. The sites and relative map positions of these proteins can be related to the gene arrangement for EBV displayed in the inverse prototype orientation (Fig. 4). For example, the HVS *EcoRI* B fragment selects a late transcript which translates to a 146-kDa protein which correlates with the late 154-kDa major capsid protein encoded by the BcLFI reading frame at a comparable position on the genome of EBV. In addition, the 115-kDa early protein translated from RNA selected with the *KpnI* F fragment of HVS corresponds to the 123-kDa DNA-binding protein encoded by BALF2 or EBV. Moreover, by using specific monoclonal antibodies (50) to identify the products of in vitro translation of hybrid-selected RNA, the immediate-early 52-kDa protein (51, 52) and the 110-kDa delayed-early DNA-binding proteins of HVS (7, 50) have been mapped to the *EcoRI* I and *KpnI* F fragments, respectively (W. Hell, E. P. Smith, H. Wolf and R. W. Honess, unpublished results). These results are consistent with the location of HVS RF8 plus RF9, RF12 plus RF13, and RF1, which identified proteins homologous to EBV BcLF1, BMLF1, and BALF2, respectively (Fig. 3, 4). The product of BMLF1 has been shown to *trans*-activate genes (11, 37) and is homologous with the HSV IE 63K protein (16). The product of BALF2 is homologous with the HSV major DNA-binding protein (48), which has roles in DNA replication and gene regulation (28).

In conclusion, parts of HVS genes encoding amino acid sequences homologous to products of EBV open reading frames have been located, and the overall gene orders of HVS thus deduced are similar to those of EBV rather than to those of the alphaherpesviruses HSV and VZV. The genetic organization of these gammaherpesviruses represents rearrangements including a circular permutation of that observed for the alphaherpesviruses VZV and HSV (Fig. 4). This gammaherpesvirus organization is also distinct from that observed for a representative of the betaherpesviruses, human cytomegalovirus (36; B. Barrell, personal communication). Furthermore, we have also noted that the DNA sequences of HVS (9) and of EBV resemble those of higher eucaryotes in having a markedly lower frequency of CpG-dinucleotides than would be expected from their mononucleotide compositions (5); there is no such deficiency in CpG-dinucleotides in the DNA sequences of the alphaherpesviruses HSV and VZV (R. W. Honess, U. A. Gompels, B. G. Barrell and R. Staden, unpublished results). The biological subclassification of these herpesviruses is therefore consistent with major features of their genetic organization.

It is of interest to note that regions of divergence within and between the gamma- and alphaherpesviruses are marked by the presence of repetitive sequences (Fig. 4). Such repeats may represent regions less subject to selective pressure or hot spots for recombination (see, for example, references 19 and 56). Thus, HVS has a loss of genetic information relative to EBV between gene blocks which are marked by repetitive families of DNA sequences in the EBV genome. The most obvious example is the lack of any

equivalent of the EBV *Bam*HI-W repeat series in the genome of HVS. Furthermore, 7 kilobases of HVS L-DNA adjacent to the left terminal repeats is hypervariable even between strains, such that similarity cannot be detected by DNA hybridization (40). Some differences in this region correlate with oncogenicity and ability to transform T cells (20, 22). Thus, although these lymphotropic herpesviruses share a similar overall organization of conserved genes, supporting their common subclassification, there are significant regions of genetic divergence that may determine elements of the distinctive biology of HVS.

LITERATURE CITED

1. Baer, R., A. T. Bankier, M. D. Biggin, P. L. Deininger, P. J. Farrell, T. J. Gibson, G. Hatful, G. S. Hudson, S. C. Satchwell, C. Seguin, P. S. Tuffnell, and B. G. Barrell. 1984. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature (London)* **310**:207-211.
2. Bankier, A. T., and B. G. Barrell. 1983. Shotgun DNA sequencing, p. 1-33. *In* R. A. Flavell (ed.), *Techniques in the life sciences*, vol. B5. Elsevier/North Holland Publishers, Ltd., Dublin.
3. Bankier, A. T., W. Dietrich, R. Baer, B. G. Barrell, F. Colbère-Garapin, B. Fleckenstein, and W. Bodemer. 1985. Terminal repetitive sequences in herpesvirus saimiri virion DNA. *J. Virol.* **55**:133-139.
4. Biggin, M. D., T. Gibson, and G. F. Hong. 1983. Buffer-gradient gels and ³⁵S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* **80**:3963-3965.
5. Bird, A. P. 1980. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* **8**:1499-1504.
6. Birnboim, H. C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**:1513-1523.
7. Blair, E. D., and R. W. Honess. 1983. DNA-binding proteins specified by herpesvirus saimiri. *J. Gen. Virol.* **64**:2697-2715.
8. Bodemer, W., H. H. Niller, N. Nitsche, B. Scholtz, and B. Fleckenstein. 1986. Organization of the thymidylate synthase gene of herpesvirus saimiri. *J. Virol.* **60**:114-123.
9. Cameron, K. R., T. Stamminger, M. Craxton, W. Bodemer, R. W. Honess, and B. Fleckenstein. 1987. The 160,000-*M_r* virion protein encoded at the right end of the herpesvirus saimiri genome is homologous to the 140,000-*M_r* membrane antigen encoded at the left end of the Epstein-Barr virus genome. *J. Virol.* **61**:2063-2070.
10. Chen, C. W., and C. A. Thomas. 1980. Recovery of DNA segments from agarose gels. *Anal. Biochem.* **101**:339-341.
11. Chevallier-Greco, A., E. Manet, P. Chavrier, C. Mosnier, J. Daillie, and A. Sargeant. 1986. Both Epstein-Barr virus (EBV)-encoded *trans*-acting factors, EB1 and EB2, are required to activate transcription from an EBV early promoter. *EMBO J.* **5**:3243-3249.
12. Costa, R. H., K. G. Draper, T. J. Kelly, and E. K. Wagner. 1985. An unusual spliced herpes simplex virus type 1 transcript with sequence homology to Epstein-Barr virus DNA. *J. Virol.* **54**:317-328.
13. Davison, A. J., and D. J. McGeoch. 1986. Evolutionary comparisons of the S segments in the genomes of herpes simplex virus type 1 and varicella-zoster virus. *J. Gen. Virol.* **67**:597-611.
14. Davison, A. J., and J. E. Scott. 1986. The complete DNA sequence of varicella-zoster virus. *J. Gen. Virol.* **67**:1759-1816.
15. Davison, A. J., and J. E. Scott. 1986. DNA sequence of the major capsid protein gene of herpes simplex virus type 1. *J. Gen. Virol.* **67**:2279-2286.
16. Davison, A. J., and P. Taylor. 1987. Genetic relations between varicella-zoster virus and Epstein-Barr virus. *J. Gen. Virol.* **68**:1067-1079.
17. Davison, A. J., and N. M. Wilkie. 1983. Location and orientation of homologous sequences in the genomes of five herpesviruses. *J. Gen. Virol.* **64**:1927-1942.
18. Dayhoff, M. O., W. C. Baker, and L. T. Hunter. 1983. Establishing homologies in protein sequences. *Methods Enzymol.* **91**:524-545.
19. de Klein, A., T. van Agthoven, C. Groffen, N. Heisterkamp, J. Groffen, and G. Grosveld. 1986. Molecular analysis of both translocation products of a Philadelphia-positive CML patient. *Nucleic Acids Res.* **14**:7071-7082.
20. Desrosiers, R. C., A. Bakker, J. Kamine, L. A. Falk, R. D. Hunt, and N. W. King. 1985. A region of the herpesvirus saimiri genome required for oncogenicity. *Science* **228**:184-187.
21. Desrosiers, R. C., C. Mulder, and B. Fleckenstein. 1979. Methylation of herpesvirus saimiri DNA in lymphoid tumor cell lines. *Proc. Natl. Acad. Sci. USA* **76**:3839-3843.
22. Desrosiers, R. C., D. P. Silva, L. M. Waldron, and N. L. Letvin. 1986. Nononcogenic deletion mutants of herpesvirus saimiri are defective for in vitro immortalization. *J. Virol.* **57**:701-705.
23. Falk, L. G., L. A. Wolfe, and F. Deinhardt. 1972. Isolation and herpesvirus saimiri from blood of squirrel monkeys (*Saimiri sciureus*). *J. Natl. Cancer Inst.* **48**:1499-1505.
24. Feng, D. F., M. S. Johnson, and R. F. Doolittle. 1984. Aligning amino acid sequences: comparisons of commonly used methods. *J. Mol. Evol.* **21**:112-125.
25. Fleckenstein, B., and R. C. Desrosiers. 1982. Herpesvirus saimiri and herpesvirus ateles, p. 253-332. *In* B. Roizman (ed.), *The herpesviruses*, vol. 1. Plenum Publishing Corp., New York.
26. Gibson, T., P. Stockwell, M. Ginsburg, and B. Barrell. 1984. Homology between two EBV early genes and the HSV ribonucleotide reductase and 38K genes. *Nucleic Acids Res.* **12**:5087-5099.
27. Goad, W. B. 1986. Computational analysis of genetic sequences. *Ann. Rev. Biophys. Biophys. Chem.* **15**:79-95.
28. Godowski, P. J., and D. M. Knipe. 1985. Identification of a herpes simplex virus function that represses late gene expression from parental viral genomes. *J. Virol.* **55**:357-365.
29. Gompels, U., and A. Minson. 1986. The properties and sequence of glycoprotein H of herpes simplex virus type 1. *Virology* **153**:230-247.
30. Hell, W., S. Modrow, and H. Wolf. 1985. Mapping of herpesvirus saimiri proteins on the viral genome: proteins dependent and not dependent on viral DNA synthesis. *J. Virol.* **56**:414-418.
31. Honess, R. W. 1984. Herpes simplex and "the herpes complex": diverse observations and a unifying hypothesis. *J. Gen. Virol.* **65**:2077-2107.
32. Honess, R. W., W. Bodemer, K. R. Cameron, H. H. Niller, B. Fleckenstein, and R. E. Randall. 1986. The A+T-rich genome of herpesvirus saimiri contains a highly conserved gene for thymidylate synthase. *Proc. Natl. Acad. Sci. USA* **83**:3604-3608.
33. Honess, R. W., and D. H. Watson. 1977. Unity and diversity in the herpesviruses. *J. Gen. Virol.* **37**:15-37.
34. Kieff, E., T. Dambaugh, W. King, M. Heller, A. Cheung, V. van Santen, M. Hummel, C. Beisel, and S. Fennwald. 1982. Biochemistry of Epstein-Barr virus, p. 105-150. *In* B. Roizman (ed.), *The herpesviruses*, vol. 1. Plenum Publishing Corp., New York.
35. Knust, E., S. Schirm, W. Dietrich, W. Bodemer, E. Kolb, and B. Fleckenstein. 1983. Cloning and herpesvirus saimiri DNA fragments representing the entire L-region of the genome. *Gene* **25**:281-289.
36. Kouzarides, T., A. T. Bankier, S. C. Satchwell, K. Weston, P. Tomlinson, and B. G. Barrell. 1987. Large-scale rearrangements of homologous regions in the genomes of HCMV and EBV. *Virology* **157**:397-413.
37. Lieberman, P. M., P. O'Hare, G. S. Hayward, and S. D. Hayward. 1986. Promiscuous *trans* activation of gene expression by an Epstein-Barr virus-encoded early nuclear protein. *J. Virol.* **60**:140-148.
38. Littler, E., J. Zuethen, A. A. McBridge, E. Trost Sorensen, K. L. Powell, J. E. Walsh-Arrand, and J. R. Arrand. 1986. Identification of an Epstein-Barr virus-coded thymidine kinase. *EMBO J.* **5**:1959-1966.
39. Maruyama, T., T. Gojobori, S. Aota, and T. Ikemura. 1986. Codon usage tabulated from the GenBank genetic sequence data. *Nucleic Acids Res.* **14**:5:r151-r197.
40. Medveczky, P., E. Szomolanyi, R. C. Desrosiers, and C. Mulder.

1984. Classification of herpesvirus saimiri into three groups based on extreme variation in a DNA region required for oncogenicity. *J. Virol.* **52**:938–944.
41. McGeoch, D. J., and A. J. Davison. 1986. DNA sequence of the herpes simplex virus type 1 gene encoding glycoprotein gH, and identification of homologues in the genomes of varicella-zoster virus and Epstein-Barr virus. *Nucleic Acids Res.* **14**:4281–4292.
42. McGeoch, D. J., A. Dolan, and M. C. Frame. 1986. DNA sequence of the region of the genome of herpes simplex virus type 1 containing the exonuclease gene and neighbouring genes. *Nucleic Acids Res.* **14**:3435–3448.
43. McLachlan, A. D. 1971. Tests for comparing related amino-acid sequences. Cytochrome C and cytochrome C551. *J. Mol. Biol.* **61**:409–424.
44. Messing, J. 1983. New M13 vectors for cloning. *Methods Enzymol.* **101**:20–78.
45. Morgan, D. G. 1977. Observations on the antigenic relationships between Epstein-Barr virus and herpesvirus saimiri. *J. Gen. Virol.* **36**:281–287.
46. Pellet, P. E., M. D. Biggin, B. Barrell, and B. Roizman. 1985. Epstein-Barr virus genome may encoded a protein showing significant amino acid and predicted secondary structure homology with glycoprotein B of herpes simplex virus 1. *J. Virol.* **56**:807–813.
47. Pellet, P. E., F. J. Jenkins, M. Ackermann, M. Sarmiento, and B. Roizman. 1986. Transcription initiation sites and nucleotide sequence of a herpes simplex virus 1 gene conserved in the Epstein-Barr virus genome and reported to affect the transport of viral glycoproteins. *J. Virol.* **60**:1134–1140.
48. Quinn, J. P., and D. J. McGeoch. 1985. DNA sequence of the region in the genome of herpes simplex virus type 1 containing the genes for DNA polymerase and the major DNA binding protein. *Nucleic Acids Res.* **13**:8143–8163.
49. Randall, R. E., R. W. Honess, and P. O'Hare. 1983. Proteins specified by herpesvirus saimiri: identification and properties of virus-specific polypeptides in productively infected cells. *J. Gen. Virol.* **64**:19–35.
50. Randall, R. E., C. Newman, and R. W. Honess. 1984. Isolation and characterization of monoclonal antibodies to structural and nonstructural herpesvirus saimiri proteins. *J. Virol.* **52**:872–883.
51. Randall, R. E., C. Newman, and R. W. Honess. 1984. A single major immediate-early virus gene product is synthesized in cells productively infected with herpesvirus saimiri. *J. Gen. Virol.* **65**:1215–1219.
52. Randall, R. E., C. Newman, and R. W. Honess. 1985. Asynchronous expression of the immediate-early protein of herpesvirus saimiri in populations of productively infected cells. *J. Gen. Virol.* **66**:2199–2213.
53. Roizman, B. 1982. The family *Herpesviridae*: general description, taxonomy, and classification, p. 1–23. In B. Roizman (ed.), *The herpesviruses*, vol. 1. Plenum Publishing Corp., New York.
54. Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe. 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* **143**:161–178.
55. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
56. Sawada, I., and C. W. Schmid. 1986. Primate evolution of the alpha-globin gene cluster and its Alu-like repeats. *J. Mol. Biol.* **192**:693–709.
57. Staden, R. 1982. An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucleic Acids Res.* **10**:2951–2961.
58. Staden, R. 1982. Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucleic Acids Res.* **10**:4731–4751.
59. Staden, R. 1984. Graphic methods to determine the function of nucleic acid sequences. *Nucleic Acids Res.* **12**:521–538.
60. Staden, R. 1984. Measurements of the effects that coding for protein has on a DNA sequence and their use for finding genes. *Nucleic Acids Res.* **12**:551–567.
61. Staden, R. 1986. The current status and portability of our sequence handling software. *Nucleic Acids Res.* **14**:217–231.
62. Staden, R., and A. D. McLachlan. 1982. Coding preference and its use in identifying protein coding regions in long DNA sequences. *Nucleic Acids Res.* **10**:141–156.
63. Stamminger, T., R. W. Honess, D. F. Young, W. Bodemer, E. D. Blair, and B. Fleckenstein. 1987. Organization of terminal reiterations in the virion DNA of herpesvirus saimiri. *J. Gen. Virol.* **68**:1049–1066.
64. Thompson, R., R. W. Honess, L. Taylor, J. Morran, and A. J. Davison. 1987. Varicella-zoster virus specifies a thymidylate synthetase. *J. Gen. Virol.* **68**:1449–1455.
65. Vogelstein, B., and D. Gillespie. 1979. Preparative and analytical purification of DNA from agarose. *Proc. Natl. Acad. Sci. USA* **76**:615–619.
66. Weston, K., and B. G. Barrell. 1986. Sequence of the short unique region, short repeats and part of the long repeats of human cytomegalovirus. *J. Mol. Biol.* **192**:177–208.
67. Wilbur, W. J., and D. J. Lipman. 1983. Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. USA* **80**:726–730.
68. Yanisch-Perron, C., I. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. *Gene* **33**:103–119.