JOURNAL OF VIROLOGY, Mar. 1994, p. 1886–1902
0022-538X/94/$04.00+0

Vol. 68, No. 3

Molecular Evolution of Herpesviruses: Genomic and Protein Sequence Comparisons

SAMUEL KARLIN,[1]* EDWARD S. MOCARSKI,[2] AND GABRIEL A. SCHACHTEL[3]

Department of Mathematics[1] and Department of Microbiology and Immunology,[2] Stanford University, Stanford, California 94305, and Division of Mathematics, Computer Science and Statistics, The University of Texas at San Antonio, San Antonio, Texas 78285[3]

Received 24 September 1993/Accepted 3 December 1993

Phylogenetic reconstruction of herpesvirus evolution is generally founded on amino acid sequence comparisons of specific proteins. These are relevant to the evolution of the specific gene (or set of genes), but the resulting phylogeny may vary depending on the particular sequence chosen for analysis (or comparison). In the first part of this report, we compare 13 herpesvirus genomes by using a new multidimensional methodology based on distance measures and partial orderings of dinucleotide relative abundances. The sequences were analyzed with respect to (i) genomic compositional extremes; (ii) total distances within and between genomes; (iii) partial orderings among genomes relative to a set of sequence standards; (iv) concordance correlations of genome distances; and (v) consistency with the alpha-, beta-, gammaherpesvirus classification. Distance assessments within individual herpesvirus genomes show each to be quite homogeneous relative to the comparisons between genomes. The gammaherpesviruses, Epstein-Barr virus (EBV), herpesvirus saimiri, and bovine herpesvirus 4 are both diverse and separate from other herpesvirus classes, whereas alpha- and betaherpesviruses overlap. The analysis revealed that the most central genome (closest to a consensus herpesvirus genome and most individual herpesvirus sequences of different classes) is that of human herpesvirus 6, suggesting that this genome is closest to a progenitor herpesvirus. The shorter DNA distances among alphaherpesviruses supports the hypothesis that the alpha class is of relatively recent ancestry. In our collection, equine herpesvirus 1 (EHV1) stands out as the most central alphaherpesvirus, suggesting it may approximate an ancestral alphaherpesvirus. Among all herpesviruses, the EBV genome is closest to human sequences. In the DNA partial orderings, the chicken sequence collection is invariably as close as or closer to all herpesvirus sequences than the human sequence collection is, which may imply that the chicken (or other avian species) is a more natural or more ancient host of herpesviruses. In the second part of this report, evolutionary relationships among the 13 herpesvirus genomes are evaluated on the basis of recent methods of amino acid alignment applied to four essential protein sequences. In this analysis, the alignment of the two betaherpesviruses (human cytomegalovirus versus human herpesvirus 6) showed lower scores compared with alignments within alphaherpesviruses (i.e., among EHV1, herpes simplex virus type 1, varicella-zoster virus, pseudorabies virus type 1 and Marek's disease virus) and within gammaherpesviruses (EBV versus herpesvirus saimiri). Comparisons within the alpha class generally produced the highest alignment scores, with EHV1 and pseudorabies type 1 prominent, whereas herpes simplex virus type 1 versus varicella-zoster virus show the least similarity among the alpha sequences. The within-alpha, beta, and gamma class sequence similarity scores are generally 50 to 100% higher than the between-class sequence similarity scores. These results suggest that the betaherpesviruses separated earlier than the formation of the gamma class and that the alpha class may be of the most recent ancestry. By our methods, evolutionary relationships derived from genomic comparisons versus protein comparisons differ to some extent. The dinucleotide relative abundance distances appear to discriminate DNA structure specificity more than sequence specificity. The evolutionary development of genes among viruses (and species) is more dependent on each individual gene.

The rapid accumulation of large numbers of DNA sequences affords challenging opportunities for studies of molecular evolution and phylogenetic relationships among organisms. For large genomic sequences, alignments of the sequences are generally not feasible and new methods are needed. A prototype example is herpesviruses. There are currently available seven distinct herpesvirus genomes sequenced in their entirety and substantial aggregate genomic sequences from six additional herpesviruses (Table 1). On the basis of biological characteristics, tissue tropism, genomic organization, and amino acid identities, the herpesviruses are classified into alpha, beta, and gamma types. Evolutionary relationships among herpesviruses have generally been based on comparisons of specific protein sequences and have been recently discussed by McGeoch (65–67), Lawrence et al. (57a), Ross and Binns (72), Griffin (39), Bennett et al. (4), Telford et al. (80, 81), and Bublot et al. (12), among others.

Numerous methods have been used in reconstructing phylogenetic relationships (trees) among groups of organisms (for recent reviews, see references 32, 38, 58, 68, 68a, and 78). In particular, there is a family of methods which relies on initial alignment of homologous DNA or protein sequences followed by tree construction based on various principles, including parsimony (28, 76, 77), distance matrices (17), maximum

* Corresponding author. Mailing address: Department of Mathematics, Stanford University, Stanford, CA 94305-2125. Phone: (415) 723-2204. Fax: (415) 725-2040. Electronic mail address: skarlin@gnomic.stanford.edu.

TABLE 1. The 15 genomes analyzed[a]

Genome	Class	Length (bp)	G+C (%)	Segment analyzed
EBV	Gamma	172,282	59.94	Complete genome
BHV4	Gamma	19,637	40.26	33 sequences
HVS	Gamma	112,930	34.51	Complete genome
HCMV	Beta	229,354	57.16	Complete genome
HHV6	Beta	59,005	41.75	9 sequences
EHV1	Alpha	150,223	56.67	Complete genome
HSV1	Alpha	152,260	68.28	Complete genome
HSV2	Alpha	59,382	68.63	19 sequences
VZV	Alpha	124,884	46.02	Complete genome
MDV	Alpha	29,259	44.54	13 sequences
PRV1	Alpha	53,496	72.66	20 sequences
BHV1	Alpha	23,030	70.67	10 sequences
IHV1		134,226	56.25	Complete genome
Chicken		1,001,390	50.27	
Human		1,410,904	50.99	

[a] Sequences from 13 herpesvirus genomes (7 complete genomes and substantial nonredundant sequences from 6 other genomes) were included in the data set. Random sequence collections exceeding 1 Mb from human and chicken genomes were included for additional comparisons.

likelihood (31), invariants, and paralinear distances (56, 57). For all of these methods, different sets of proteins for the same set of organisms can produce different phylogenetic reconstructions.

This study sets forth and applies a new methodology for assessing sequence evolutionary relationships among herpesvirus genomes based on distance measures, partial orderings, and concordance correlations of dinucleotide relative abundances (see Materials and Methods for precise definitions and rationale). Evolutionary distances can be assessed by using di-, tri-, or tetranucleotide relative abundance distances based on representative DNA sequences available for each organism. We further calculate distances between different subsets of sequences from the same organism, e.g., between the U_L (unique long) versus U_S (unique short) sequences of the alphaherpesviruses (α-HV). A further elaboration of our analysis yields groupings of organisms or viruses; e.g., here we construct sets of consensus α-, β-, and γ-HV and calculate for such groups an averaged set of dinucleotide relative abundances (see Materials and Methods), which helps us in evaluating the validity and consistency of the α-, β-, and γ-HV classifications.

The following questions are of interest with respect to herpesvirus DNA sequence comparisons. Is there a single alpha and gamma class? To what extent are the various human herpesviruses similar to or different from the human host or to other mammalian or avian hosts? Which herpesvirus genomes have the most (least) bias in dinucleotide relative abundances? To what extent is the U_L region similar to the U_S region within and between α-HV genomes and human cytomegalovirus (HCMV)? Is the as yet unclassified channel catfish herpesvirus (ictalurid herpesvirus 1 [IHV1]) closer to the consensus α-, β-, or γ-HV? How far are the putatively unrelated adenovirus and vaccinia virus from the various herpesviruses?

In the first part of this report, herpesvirus genomic relationships are analyzed in four ways. (i) Relative abundances of di-, tri-, and tetranucleotides are compiled, and compositional extremes are compared. (ii) Dinucleotide relative abundance distances are determined between each pair of genomes. (iii) The relative abundances of the 16 dinucleotides are used to construct partial orderings of the genomes in a multidimensional analysis. (iv) The relatedness of each pair of genomes is assessed via a measure of concordance with respect to ordered distances to the other genomes. In the second part of this report, we applied protein sequence alignment methods from three recently developed amino acid (matching) scoring regimes. The protein sequences compared are DNA polymerase, major capsid protein, a DNA-packaging (tegument) protein, and glycoprotein B (gB) available in the genomes indicated in Table 7.

TABLE 2. Short oligonucleotide relative abundance extremes[a]

Nucleotides	Human	Chicken	EBV	BHV4	HVS	HCMV	HHV6	IHV1	EHV1	HSV1	HSV2	VZV	MDV	PRV1	BHV1	Adenovirus
CG	0.42	0.50	0.60	0.18	0.33	(1.19)										0.79
TA	0.63	0.64	0.75	0.74	(0.83)	0.80	0.78	0.69		0.78	0.80			0.58		
GC				0.80				0.71								
CA/TG	1.22	1.23	1.23	1.32	1.23											
CC/GG			1.21	1.31								(1.18)				
GC					1.28										1.25	
GA/TC								1.27						1.23		
AA/TT										1.26	1.26				1.31	1.21
CGA/TCG				0.77	0.79											
TAG/CTA								0.71		0.67	0.77			0.74	0.79	
TAA/TTA										1.22				1.20		
TAC/GTA														1.24	1.34	
AAA/TTT														1.33	1.22	
CGAC/GTCG				0.60	0.78											
CTAG	(0.82)		(0.81)					0.68		0.79	0.79			0.51		0.72
ATTC/GAAT														0.55	0.77	
CCTA/TAGG														0.79	0.74	
TCGA				1.31										1.24		

[a] The high and low relative abundance values for di- and trinucleotides (see text for definition) and for tetranucleotides (14) are displayed with respect to each genome. Guided by contingency table tests, the cutoff points were set to 0.80 and 1.20. The table contains a few values, in parentheses, which do not qualify but are close to a cutoff point. Values for tri- and tetranucleotides that were low or high in only one genome are listed separately as follows.

Unique tri- and tetranucleotide underrepresentations: BHV4, AACG/CGTT 0.67; HVS, ACGC/GCGT 0.76; PRV1, ATAG/CTAT 0.64; CTTA/TAAG 0.66, ATAC/GTAT 0.70, CAAG/CTTG 0.73, ATTA/TAAT 0.75, GAAA/TTTC 0.77, GATA/TATC 0.77, CAGA/TCTG 0.79; BHV1, GTTA/TAAC 0.59; adenovirus, TCGA 0.77; vaccinia virus, CCC/GGG 0.73. Unique tri- and tetranucleotide overrepresentations: BHV4, GCG/CGC 1.21, CGCG 1.24, GACG/CGTC 1.22, CGGT/ACCG 1.21; HVS, CCC/GGG 1.29, CCGG 1.28, CGCC/GGCG 1.28, ACGT 1.25, CTCG/CGAG 1.21; PRV1, TTAG/CTAA 1.40, CAAT/ATTG 1.30, ATAT 1.30, TGAC/GTCA 1.28, TTCC/GGAA 1.26, TCGA 1.24; BHV1, TAGC/GCTA 1.24; vaccinia virus, CCA/TTG 1.23.

MATERIALS AND METHODS

Genomic comparisons. (i) Symmetrized frequencies and relative abundances. Let f_X denote the frequency of the nucleotide X (A, C, G, or T) in the sequence at hand, f_{XY} the frequency of dinucleotide XY, and so on. A standard assessment of dinucleotide bias is through the odds ratio: $\rho_{XY} = f_{XY}/f_X f_Y$. For ρ_{XY} at least 1.20 or at most 0.80 (see Table 2, footnote *a*), the doublet XY is considered to be of high or low relative abundance, respectively, compared with a random association of mononucleotides. The formula has to be modified to accommodate the complementary antiparallel structure of double-stranded DNA. This can be accomplished by considering the union of the given DNA sequence S and its inverted complement sequence S^T into $S + S^T = S^*$. In S^*, the frequency f_A^* of the mononucleotide A is $f_A^* = f_T^* = (f_A + f_T)/2$ and $f_C^* = f_G^* = (f_C + f_G)/2$, where f_A, f_T, f_C, and f_G are the mononucleotide frequencies in S. Similarly, $f_{GT}^* = (f_{GT} + f_{AC})/2$ is the symmetrized double stranded frequency of GT/AC, etc. For our purposes, the dinucleotide odds ratio measure that accounts for the complementary antiparallel structure of double-stranded DNA is taken to be $\rho_{GT}^* = f_{GT}^*/f_G^* f_T^* = 2(f_{GT} + f_{AC})/(f_G + f_C)(f_T + f_A)$ and similarly for all other dinucleotides. The deviation of ρ_{GT}^* from 1 can be construed as a measure of dinucleotide bias of GT/AC (14). A corresponding third-order measure of similar type is $\gamma_{XYZ}^* = f_{XYZ}^* f_X^* f_Y^* f_Z^* / f_{XY}^* f_{YZ}^* f_{XNZ}^*$, where N is any nucleotide. Higher-order measures based on tetranucleotide (or longer oligonucleotide) frequencies and criteria for extremes are also available.

(ii) Dinucleotide relative abundance distance measures. Let $\rho_{ij}^* = f_{ij}^*/f_i^* f_j^*$ be the symmetrized relative abundance measure of the dinucleotide (i,j). We propose a measure of dinucleotide distance between two sequences f and g (e.g., sequences from different organisms or from different regions of a genomic sequence), which controls for differing base compositions; the dinucleotide relative abundance distance, $\delta(f,g)$, is defined as

$$\delta(f,g) = \sum | \rho_{ij}^*(f) - \rho_{ij}^*(g) | w_{ij} \qquad (1)$$

where the sum extends over all dinucleotides and $w_{ij} = 1/16$ or some other natural weights. This distance is in essence an absolute difference of the second-order residuals: $(f_{ij}^*/f_i^* f_j^* - 1)$ and $(g_{ij}^*/g_i^* g_j^* - 1)$. The distance defined by equation 1 is to be contrasted to the dinucleotide frequency distance,

$$d(f,g) = \sum | f_{ij}^* - g_{ij}^* | w_{ij} \qquad (2)$$

The absolute frequency metric $d(f,g)$ of equation 2, even for related species, tends to yield poor comparative distance assessments because of mononucleotide compositional biases. Specifically, the dinucleotide frequency distances (equation 2) among the human herpesviruses with weights $w_{ij} = 1/16$ yield

f = HSV1 versus g	EBV	HCMV	VZV	human
$d(f,g)$	0.016	0.019	0.030	0.032

which parallel the genomic G+C contents (herpes simplex virus type 1 [HSV1; 68%], Epstein-Barr virus [EBV; 60%], HCMV [57%], varicella-zoster virus [VZV; 46%], human [about 40%]), placing VZV farthest among the three herpesviruses from HSV1.

In contrast, the dinucleotide relative abundance distance assessments (equation 1) are quite compelling in relating the herpesvirus sequences with each other and to the human host (for complete comparisons, see Table 3):

f = HSV1 versus g	VZV	HCMV	EBV	human
$\delta(f,g)$	0.081	0.090	0.160	0.209

affirming VZV as the closest to HSV1 among VZV, HCMV, and EBV despite their very different genomic G+C contents but consistent with their classifications as neurotropic viruses with similar genomic organizations.

(iii) Examples of dinucleotide relative abundance distances among eukaryotic genomes. To help in the interpretations of the relative abundance distances, we give examples for various levels of distances. The sequences of these examples generally range from 200 kb up to 1 or 2 Mb: closely related ($\delta \leq 0.040$), δ(cow, pig) = 0.015, δ(human, cow) = 0.033, and δ(*Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*) = 0.034; moderately related (δ range, 0.050 to 0.075), δ(*Drosophila melanogaster* and *Bombyx mori*) = 0.058 and δ(*S. cerevisiae*, *Neurospora crassa*) = 0.048; weakly related (δ range, 0.080 to 0.110), δ(human, *S. cerevisiae*) = 0.107, δ(human, trout) = 0.090, and δ(*N. crassa*, chicken) = 0.114; distantly related (δ range, 0.120 to 0.150), δ(human, *D. melanogaster*) = 0.145 and δ(human, *Caenorhabditis elegans*) = 0.150; distant (δ range, 0.160 to 0.190), δ(mouse, *C. elegans*) = 0.177 and δ(pig, *B. mori*) = 0.186; and very distant ($\delta \geq 0.200$), δ(human, *Escherichia coli*) = 0.210 and δ(human, *Bacillus subtilis*) = 0.279.

(iv) Consensus virus sequences. Consensus α-, β-, and γ-HV sequences are generated by appropriately combining the available sequences: α-HV includes equine herpesvirus 1 (EHV1), HSV1, HSV2, VZV, Marek's disease virus (MDV), pseudorabies virus type 1 (PRV1), and bovine herpesvirus 1 (BHV1), β-HV includes HCMV and human herpesvirus 6 (HHV6), and γ-HV includes EBV, herpesvirus saimiri (HVS), and BHV4. A global consensus herpesvirus sequence combines all of the foregoing sequences. The dinucleotide relative abundances for the consensus viruses are determined by weighting each component genome equally. To this end, we average the symmetrized mono- and dinucleotide frequencies over the group members and then calculate the corresponding ρ_{XY}^* value. For example,

$$f_{XY}^*(\gamma) = \frac{1}{3}\left[f_{XY}^*(\text{EBV}) + f_{XY}^*(\text{HVS}) + f_{XY}^*(\text{BHV4})\right];$$

$$\rho_{XY}^*(\gamma) = \frac{f_{XY}^*(\gamma)}{f_X^*(\gamma) f_Y^*(\gamma)}$$

(v) Kendall tau correlations of δ-distance orderings. For each sequence standard s, we calculate the dinucleotide relative abundance distances $\delta(g,s)$ (equation 1) to each other sequence, g (Table 3). For two different standards, s and t, we determine the corresponding δ-distance orderings $\delta(g,s)$ and $\delta(g,t)$, excluding $g = t$ from the first array and $g = s$ from the second array. The Kendall tau correlation $\tau(s,t)$ (see below) assesses the degree of concordance between the two orderings.

Let $\delta_1(s) = \delta(g_1,s)$, $\delta_2(s)$, ..., $\delta_n(s)$ denote the distances of n distinct sequences (genomes) g_1, g_2, ..., g_n from the standard s, and $\delta_1(t) = \delta(g_1,t)$, and let $\delta_2(t)$, ..., $\delta_n(t)$ denote the corresponding distances from the standard t (excluding s and t from both). For each pair of genomes i and j ($1 \leq i,j \leq n$), we set

TABLE 3. Distance arrays for 20 standards[a]

Human		Chicken		Gamma		EBV		BHV4		HVS		Beta		HCMV		HHV6		Unit	
Human	0.000	Chicken	0.000	Gamma	0.000	EBV	0.000	BHV4	0.000	HVS	0.000	Beta	0.000	HCMV	0.000	HHV6	0.000	Unit	0.000
Chicken	0.037	Human	0.037	EBV	0.064	Gamma	0.064	Gamma	0.131	Chicken	0.126	HCMV	0.038	Beta	0.038	Beta	0.044	EHV1	0.049
EBV	0.071	EBV	0.068	Human	0.076	Chicken	0.068	EBV	0.147	Gamma	0.130	HHV6	0.044	HHV6	0.067	EHV1	0.059	MDV	0.076
Gamma	0.076	Gamma	0.079	Chicken	0.079	Human	0.071	Human	0.154	Human	0.131	EHV1	0.061	EHV1	0.081	HSV1	0.063	HHV6	0.078
HVS	0.131	HVS	0.126	EHV1	0.130	MDV	0.130	Chicken	0.172	EBV	0.155	HSV1	0.063	HSV1	0.090	HSV2	0.064	Beta	0.080
BHV4	0.154	EHV1	0.156	HVS	0.130	EHV1	0.144	HVS	0.176	EHV1	0.174	HSV2	0.072	Unit	0.100	MDV	0.065	HSV1	0.085
MDV	0.175	Virus	0.167	BHV4	0.131	BHV4	0.147	MDV	0.216	BHV4	0.176	Virus	0.072	Virus	0.101	HCMV	0.067	Virus	0.085
HHV6	0.185	HHV6	0.168	MDV	0.140	HHV6	0.147	Unit	0.226	HCMV	0.184	Alpha	0.075	HSV2	0.104	Virus	0.069	PRV1	0.088
EHV1	0.188	MDV	0.169	Virus	0.141	Unit	0.150	EHV1	0.227	Unit	0.193	Unit	0.080	Alpha	0.110	Alpha	0.078	Alpha	0.095
Unit	0.201	BHV4	0.172	HHV6	0.157	HVS	0.155	HHV6	0.228	BHV1	0.194	MDV	0.096	BHV1	0.112	Unit	0.078	HCMV	0.100
Virus	0.201	Unit	0.180	Unit	0.158	Virus	0.156	HSV1	0.241	Beta	0.196	VZV	0.099	MDV	0.114	PRV1	0.093	HSV2	0.102
HSV1	0.209	Beta	0.185	HSV1	0.173	HSV1	0.160	Virus	0.246	HHV6	0.213	BHV1	0.110	VZV	0.124	VZV	0.105	VZV	0.116
PRV1	0.210	HSV1	0.186	Beta	0.176	Beta	0.176	Beta	0.257	Virus	0.217	PRV1	0.111	PRV1	0.137	IHV1	0.112	BHV1	0.118
Beta	0.214	HCMV	0.190	Alpha	0.177	HSV2	0.179	PRV1	0.257	MDV	0.227	IHV1	0.131	IHV1	0.147	BHV1	0.132	IHV1	0.131
HCMV	0.224	PRV1	0.198	VZV	0.179	HCMV	0.183	HSV2	0.260	HSV1	0.234	EBV	0.176	EBV	0.183	EBV	0.147	EBV	0.150
HSV2	0.225	Alpha	0.200	BHV1	0.183	PRV1	0.183	VZV	0.268	PRV1	0.234	Gamma	0.176	HVS	0.184	Gamma	0.157	Gamma	0.158
Alpha	0.232	HSV2	0.207	HCMV	0.186	Alpha	0.188	HCMV	0.271	Alpha	0.241	Chicken	0.185	Gamma	0.186	Chicken	0.168	Chicken	0.180
VZV	0.238	BHV1	0.210	HSV2	0.193	VZV	0.190	Alpha	0.272	HSV2	0.252	HVS	0.196	Chicken	0.190	Human	0.185	HVS	0.193
BHV1	0.245	VZV	0.230	PRV1	0.206	BHV1	0.201	IHV1	0.282	VZV	0.264	Human	0.214	Human	0.224	HVS	0.213	Human	0.201
IHV1	0.267	IHV1	0.255	IHV1	0.254	IHV1	0.222	BHV1	0.304	IHV1	0.301	BHV4	0.257	BHV4	0.271	BHV4	0.228	BHV4	0.226

Alpha		EHV1		HSV1		HSV2		VZV		MDV		PRV1		BHV1		IHV1		Virus	
Alpha	0.000	EHV1	0.000	HSV1	0.000	HSV2	0.000	VZV	0.000	MDV	0.000	PRV1	0.000	BHV1	0.000	IHV1	0.000	Virus	0.000
Virus	0.037	Virus	0.048	HSV2	0.023	HSV1	0.023	HSV1	0.081	Virus	0.061	Unit	0.088	EHV1	0.096	PRV1	0.103	Alpha	0.037
HSV2	0.058	Unit	0.049	Virus	0.054	Alpha	0.058	Virus	0.083	HHV6	0.065	HHV6	0.093	Virus	0.100	HSV2	0.107	EHV1	0.048
HSV1	0.059	HHV6	0.059	Alpha	0.059	HHV6	0.064	Alpha	0.088	EHV1	0.076	IHV1	0.103	Alpha	0.104	HHV6	0.112	HSV1	0.054
EHV1	0.066	Beta	0.061	Beta	0.063	Virus	0.070	HSV2	0.094	Unit	0.076	Beta	0.111	Beta	0.110	HSV1	0.112	MDV	0.061
Beta	0.075	Alpha	0.066	HHV6	0.063	Beta	0.072	EHV1	0.095	Alpha	0.080	HSV2	0.115	HCMV	0.112	MDV	0.127	HHV6	0.069
HHV6	0.078	MDV	0.076	EHV1	0.076	MDV	0.089	Beta	0.099	HSV1	0.086	EHV1	0.116	Unit	0.118	Beta	0.131	HSV2	0.070
MDV	0.080	HSV1	0.076	VZV	0.081	EHV1	0.094	MDV	0.103	HSV2	0.089	MDV	0.129	HHV6	0.132	Unit	0.131	Beta	0.072
VZV	0.088	HCMV	0.081	Unit	0.085	VZV	0.094	HHV6	0.105	Beta	0.096	HSV1	0.129	HSV1	0.138	Alpha	0.140	VZV	0.083
Unit	0.095	HSV2	0.094	MDV	0.086	Unit	0.102	Unit	0.116	VZV	0.103	HCMV	0.137	HSV2	0.150	HCMV	0.147	Unit	0.085
BHV1	0.104	VZV	0.095	HCMV	0.090	HCMV	0.104	HCMV	0.124	HCMV	0.114	Alpha	0.145	VZV	0.154	VZV	0.149	BHV1	0.100
HCMV	0.110	BHV1	0.096	IHV1	0.112	IHV1	0.107	IHV1	0.149	IHV1	0.127	Virus	0.146	MDV	0.156	Virus	0.152	HCMV	0.101
IHV1	0.140	PRV1	0.116	PRV1	0.129	PRV1	0.115	BHV1	0.154	PRV1	0.129	BHV1	0.167	PRV1	0.167	EHV1	0.156	Gamma	0.141
PRV1	0.145	Gamma	0.130	BHV1	0.138	BHV1	0.150	Gamma	0.179	EBV	0.130	VZV	0.181	Gamma	0.183	EBV	0.222	PRV1	0.146
Gamma	0.177	EBV	0.144	EBV	0.160	EBV	0.179	PRV1	0.181	Gamma	0.140	EBV	0.183	HVS	0.194	BHV1	0.230	IHV1	0.152
EBV	0.188	Chicken	0.156	Gamma	0.173	Gamma	0.193	EBV	0.190	BHV1	0.156	Chicken	0.198	EBV	0.201	Gamma	0.254	EBV	0.156
Chicken	0.200	IHV1	0.156	Chicken	0.186	Chicken	0.207	Chicken	0.230	Chicken	0.169	Gamma	0.206	Chicken	0.210	Chicken	0.255	Chicken	0.167
Human	0.232	HVS	0.174	Human	0.209	Human	0.225	Human	0.238	Human	0.175	Human	0.210	IHV1	0.230	Human	0.267	Human	0.201
HVS	0.241	Human	0.188	HVS	0.234	HVS	0.252	HVS	0.264	BHV4	0.216	HVS	0.234	Human	0.245	BHV4	0.282	HVS	0.217
BHV4	0.272	BHV4	0.227	BHV4	0.241	BHV4	0.260	BHV4	0.268	HVS	0.227	BHV4	0.257	BHV4	0.304	HVS	0.301	BHV4	0.246

[a] Distances $\delta(fg)$ from various standards (equation 1), i.e., from each of the individual genomes, the three group consensus sequences, (alpha class [EHV1, HSV1, HSV2, VZV, MDV, PRV1, and BHV1], beta class [HCMV and HHV6], and gamma class [EBV, BHV4, and HVS]), the global consensus virus sequence (see text), and the unit (all $\rho_{ij}^{*} \equiv 1$).

$$\tau_{ij} = \begin{cases} +1 \text{ if } [\delta_i(s) - \delta_j(s)]\,[\delta_i(t) - \delta_j(t)] > 0 \\ -1 \text{ if } [\delta_i(s) - \delta_j(s)]\,[\delta_i(t) - \delta_j(t)] < 0 \\ 0 \text{ otherwise} \end{cases}$$

The Kendall tau correlation coefficient of the distance arrays for standards s and t is defined by

$$\tau(s,t) = \frac{\sum \tau_{ij}}{\sqrt{n(n-1) - 2a}\,\sqrt{n(n-1) - 2b}}$$

where a (b) is the number of ties among pairs of distances $\delta_i(s)$ and $\delta_j(s)$ ($\delta_i(t)$ and $\delta_j(t)$). Clearly, $-1 \leq \tau(s,t) \leq +1$, with equality holding only when the distance arrays of s and t are completely concordant [$\tau(s,t) = 1$] or completely discordant [$\tau(s,t) = -1$] (8, 42).

Score-based methods of sequence comparisons. For amino acid sequence comparison, a wide range of scoring regimes have been proposed (10, 25, 41). The PAM, BLOSUM, and SISS scoring matrices implemented via the computer program BLASTP (2) will identify high scoring segments common to different protein sequences. The PAM similarity scores have been developed from considerations of evolutionary amino acid replacements in homologous genes from different species (25, 48). The BLOSUM score matrices were constructed by centering on blocks of functional motifs from various functional and structural protein classes (41). The SISS scoring scheme is based on screening of statistically significant long segments among protein sequences (10). All segment pairs with scores significant at the 1% level (i.e., those with probability less than 0.01 of attaining a score of at least as high for a segment pair in random sequences of the same lengths and amino acid frequencies) can be identified. One way of scoring global similarity between two protein sequences is as follows. For each pair of protein sequences, the significant average similarity (SAS) score is the maximal value with respect to all sets of consistently ordered significantly scoring segments (overlaps are eliminated), calculated by summing these segment scores and dividing by the minimal length of the two protein sequences.

How should one interpret SAS scores? Since the scores for amino acid identities with the PAM120 matrix average 5.3, a SAS score of 2.00 generally reflects about 30 to 40% identity; a SAS score of 3.00 corresponds to about 50 to 60% identity, and a SAS score exceeding 4.00 carries at least 75% identity.

RESULTS

Compositional biases of short oligonucleotides in herpesvirus genomes. The objective of this section is to compare compositional extremes for all di-, tri-, and tetranucleotide relative abundances across the different herpesvirus genomes of Table 1. The criteria for significantly high and low relative abundances are described in Table 2, footnote *a*. The following characteristics stand out.

(i) Extremes for γ-HV. All γ-HV feature low relative abundances of the dinucleotides CG and TA (CpG and TpA suppression), a characteristic that this herpesvirus group shares with human and chicken sequences. The TA dinucleotides also occur with borderline low relative abundance in β-HV sequences and in several α-HV sequences, with ρ^*_{TA} values consistently less than 1 in all α-HV examined.

All γ-HV are CpG suppressed (Table 2), while CpG generally shows normal relative abundance in α- and β-HV genomes. Reasons for these differences are unknown. MDV and HHV6, which possess some biological features of γ-HV (lymphotropism), do not show any of the extreme dinucleotide relative abundances evident in the other groups (Table 2). CpG suppression is noted for a 5-kb stretch covering the immediate-early gene segments in HCMV, murine cytomegalovirus, simian cytomegalovirus, and HHV6 (62). It is intriguing that globally, the most overrepresented dinucleotide in HCMV is CpG (ρ^*_{CG} = 1.19). The large range in CpG frequency among alpha-, beta-, and gamma-HV is surprising, but only the gamma-HV show CpG suppression characteristic of animal genomes.

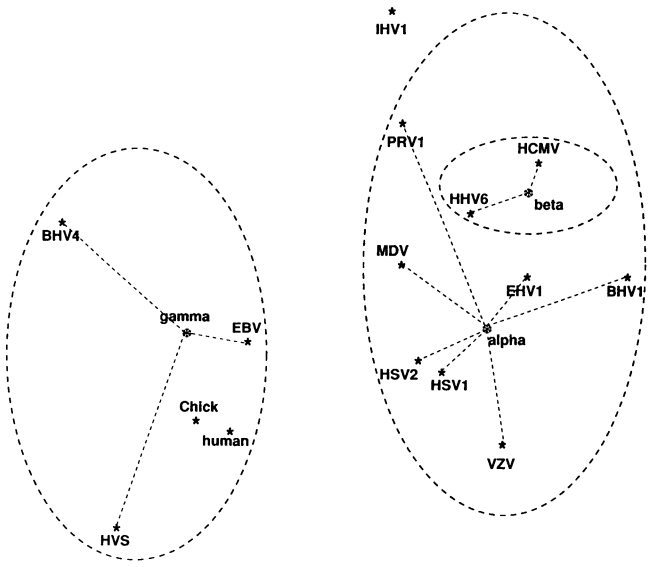

FIG. 1. Two-dimensional schematic distance representation of herpesviruses based on relative dinucleotide abundance distances (equation 1). Although mapping the high-dimensional distance space for all pairs of herpesviruses into two dimensions causes considerable reduction of information, the picture still reflects major features of the relative distances among individual herpesviruses and herpesvirus classes. (See also the footnote to Table 3.)

The most overrepresented dinucleotide among the gamma-HV is CA/TG. For vertebrate sequences, the methylase/deamination/mutation scenario might produce an excess of CA/TG doublets from CG (6, 79). Although it is unknown whether CG dinucleotides are methylated during replication, various degrees of methylation of the EBV genome in different latently infected cell lines, ranging from unmethylated to an extensively methylated state, have been observed. This methylation has been correlated with differential EBV latent gene expression, depending on the degree of methylation (46, 59, 63).

(ii) Compositional biases in β-HV genomes. Apart from the marginally low dinucleotide TA, both β-HV sequences analyzed, HHV6 and HCMV, show no significantly high or low relative abundances among di-, tri-, or tetranucleotides.

(iii) Compositional variation among α-HV. EHV1, MDV, and VZV possess normal relative abundances of all short oligonucleotides. For example, the dinucleotide ρ^*_{XY} values (XY is any dinucleotide) for EHV1 range from 0.88 to 1.09. The lowest ρ^*_{XY} value in MDV is ρ^*_{TA} = 0.90.

The trinucleotide TAG/CTA, of low relative abundance in many eukaryotes and prokaryotes (14), is of significantly low abundance in several α-HV sequences (the stop codon TAA is not significantly low [14]). The tetranucleotide CTAG, extremely rare in all bacterial species studied to date (14, 51, 64),

is also of low relative abundance in most α-HV sequences. PRV1 carries the most extreme examples with respect to over- and underabundant tetranucleotides. Relative to extremes of short oligonucleotide relative abundances, the channel catfish virus (IHV1) is most similar to an α-HV.

Total distance genomic orderings. Table 3 presents all distances $\delta(f,g)$ between individual sequences and the consensus groups or from the unit standard ($\rho^*_{ij} = 1$ for all i,j). See also the pictorial depiction of distance representations (Fig. 1).

(i) Diversity among γ-HV. The large distances among the γ-HV (EBV, HVS, and BHV4) suggests that this is a diverse group; however, each member is invariably more distant from the individual α- and β-HV than from other γ-HV. The closest herpesvirus genome to BHV4 is EBV.

BHV4 and HVS are farthest or second farthest from all non γ-HV sequences. Although these are like the other γ-HV sequences and are closest to human and chicken sequences, the distances are high (Table 3). The greatest distance, $\delta = 0.304$, among all herpesviruses is observed between the two bovine herpesviruses, BHV4 and BHV1, and the second-farthest distance ($\delta = 0.301$) is between HVS and IHV1. IHV1 is closer to (although still quite far from) the β-HV group than to the γ- or α-HV group. Among herpesviruses, γ-HV sequences generally are closest to the chicken and human sequences; of these, EBV is the closer by a factor of 2.

(ii) Is there a single α-HV class? On the DNA level, the α-HV, EHV1, HSV1, HSV2, VZV, and perhaps MDV form a relatively close group (Fig. 1), while BHV1 and PRV1 lie toward the edge of this grouping but still closer to α-HV than to γ-HV. The distance from PRV1 to BHV1 is about the same as the distance of EBV to HVS. The subgroupings among α-HV are different when comparing protein sequences (see below).

The two closest viruses are HSV1 and HSV2 [δ(HSV1, HSV2) = 0.023], and the second-closest pair is HHV6 and EHV1 ($\delta = 0.059$ of moderate distance; see Materials and Methods). The consensus α-HV sequence is close to the global herpesvirus sequence ($\delta = 0.037$) at about the distance of cow with human ($\delta = 0.035$). Interestingly, and possibly related to its biology, of all of the α-HV, MDV is closest to the γ-HV sequences. The EBV genome tends to be the closest of all γ-HV to the other herpesvirus classes.

(iii) How distinctive is the β-HV class? The β-HV (especially HHV6) and the α-HV are largely interspersed (Table 3 and Fig. 1) and are about equally distant from the γ-HV sequences. Broadly, the β-HV and the more central α-HV sequences (EHV1, MDV, HSV1, HSV2, and VZV) are close to weakly distant, contrasted to the γ-HV, for which all mutual distances are far (see Materials and Methods). Moreover, HHV6 is closest of all individual genomes to the consensus α-HV, consensus β-HV, and consensus global herpesvirus standards.

(iv) Randomness of herpesvirus sequences. The most random herpesvirus genome, g, [in the sense of having ρ^*_{ij} values generally close to 1, measured by the distance $\delta(1,g)$] is EHV1 (0.049), followed by MDV (0.076) and HHV6 (0.078). Consistent with these assessments, EHV1 shows no significantly high or low relative abundances in di-, tri-, and tetranucleotide frequencies. The α-HV genomes in general tend to be more random, i.e., closer to the unit standard, than do the members of the γ-HV class (or do the human and chicken sequences).

(v) Herpesvirus distances to adenovirus and vaccinia virus (data not shown). It is useful (in part serving as a control) to compare the herpesviruses with other relatively large animal virus genomes, e.g., those of vaccinia virus (191,737 bp, G+C = 33.5%) and adenovirus (35,937 bp, G+C = 55.2%). For adenovirus, we observe a single borderline underabundant dinucleotide ($\rho^*_{TA} = 0.79$) and a single borderline overabundant dinucleotide TT/AA ($\rho^*_{TT/AA} = 1.21$). This virus does not possess significant trinucleotide extremes. There are two significantly underabundant tetranucleotides, $\tau^*_{CTAG} = 0.72$ and $\tau^*_{TCGA} = 0.77$ (interestingly both 4-bp palindromes), and no overabundant tetranucleotides. The vaccinia virus genome has no significant extreme dinucleotide or tetranucleotide relative abundances.

In comparing δ distances (equation 1) of the adenovirus genome with herpesvirus sequences, we find δ(adenovirus, EBV) = 0.147, δ(adenovirus, HVS) = 0.180, and δ(adenovirus, BHV4) = 0.221, generally distances comparable to those among the γ-HV sequences. The distances to the α- and β-HV reflect weak relatedness.

Vaccinia virus is not particularly close to any herpesvirus, but in δ distances, δ(vaccinia virus, HHV6) = 0.089 and δ(vaccinia virus, MDV) = 0.096, which correspond to weak relatedness; the vaccinia virus-to-γ-HV distances are very distant (≥0.206) with respect to protein comparisons (27) (see Table 9).

Evolution and homogeneity assessments of U_S versus U_L sequences. Representatives of the α-HV are often neurotropic and have similar genome organizations. Their genomes are composed of two unique sequences, U_L and U_S, each flanked by inverted repeats, TR_L/IR_L and IR_S/TR_S, respectively, often with U_L and U_S of different sizes. The β-HV display varied genomic architecture: HCMV strains contain U_L and U_S with four isomeric forms paralleling those of HSV1, whereas HHV6 carries only terminal repeats in direct orientation. The γ-HV types vary widely in their genomic layouts.

The U_S regions among α-HV differ in numbers of open reading frames, orientation, and order, and they show relatively little amino acid sequence similarity. Davison and McGeoch (24) and McGeoch (65, 66) proposed that many U_S genes of α-HV have been generated from gene duplication, expansion, and contraction or by independent evolution. There are stretches of gene colinearity (in direct and/or in inverted order) of the U_L section. About 40 genes in the U_L region of HSV1 are reasonably conserved between α-, β-, and γ-HV genomes.

The gene layout in the S component of α-HV varies widely.

TABLE 4. Within-genome U_L versus U_S distances in four completely sequenced herpesviruses

Virus	U_L		U_S		Length ratio, U_S/U_L (%)	Distance, $\delta(U_S, U_L)$
	Length (bp)	G+C (%)	Length (bp)	G+C (%)		
EHV1	112,870	55	11,861	52	11	0.061
HCMV	166,949	58	35,418	56	21	0.032
HSV1	107,943	67	12,979	64	12	0.066
VZV	104,836	44	5,232	43	5	0.036

TABLE 5. Between-genome distances of U_L versus U_S

Comparison	$\delta(U_L, U_L)$	$\delta(U_S, U_S)$	$\delta(U_L, U_S)$	$\delta(U_S, U_L)$
HSV1, EHV1	0.079	0.105	0.087	0.118
HSV1, HCMV	0.079	0.107	0.076	0.112
HSV1, VZV	0.080	0.087	0.115	0.076
VZV, EHV1	0.096	0.119	0.115	0.128
VZV, HCMV	0.123	0.125	0.118	0.144
EHV1, HCMV	0.074	0.074	0.084	0.091

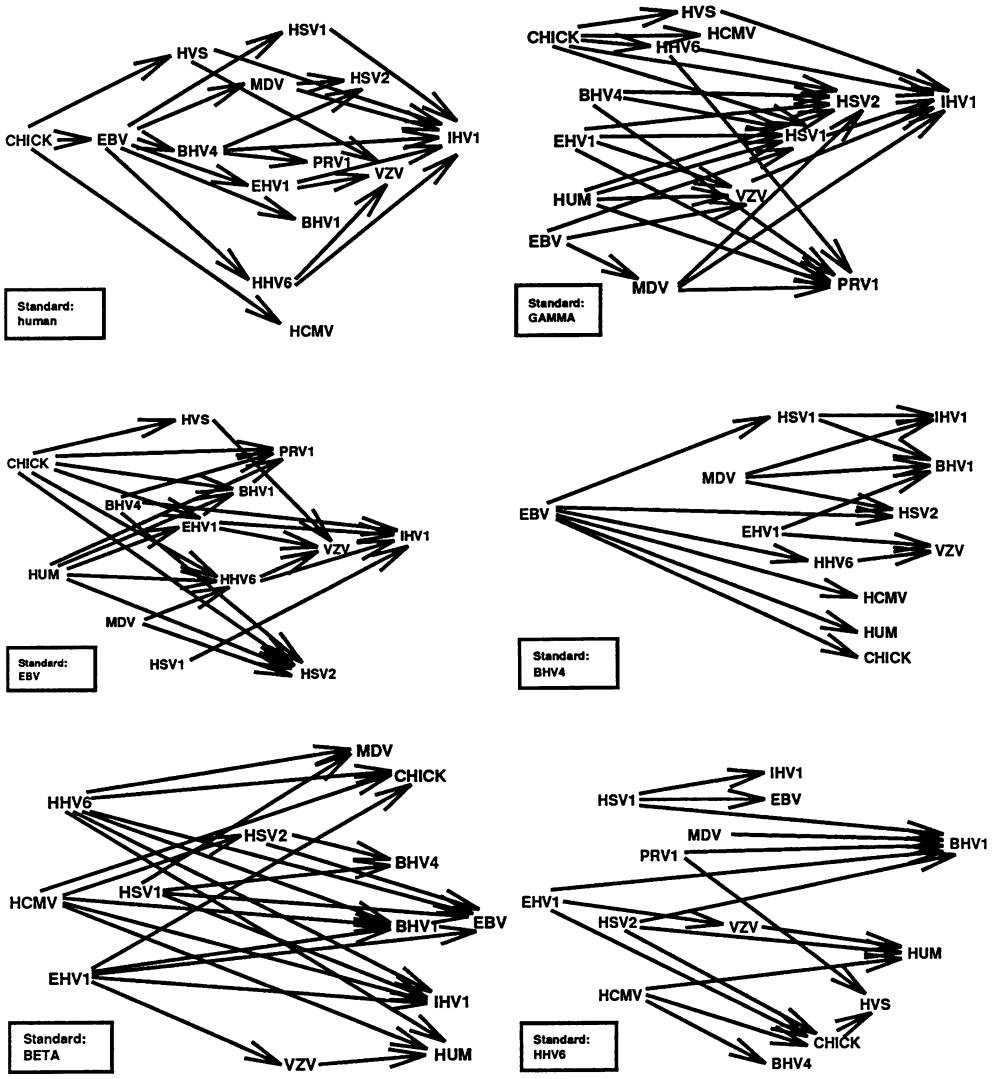

FIG. 2. Partial orderings for 12 standards. The left-to-right positions of the genomes reflect on their relative closeness to the standard (see text). An arrow from genome A to genome B indicates that A dominates B with respect to the given standard (see Materials and Methods). HUM, human.

In particular, the U_S component length excluding the inverted repeats varies from about 5 kb in VZV to about 13 kb in HSV1, up to 35 kb in the β-HV example, HCMV. The dinucleotide relative abundance distances were determined for the sequences of U_L versus U_S of the four (HSV1, VZV, EHV1, and HCMV) complete genomes. Inspection of Table 4 reveals that the U_L and U_S components within HCMV and within VZV are closer by a factor of 2 compared to the U_S versus U_L components within HSV1 and EHV1, with the $\delta(U_S, U_L)$ distances for these four genomes ranging from very close (0.032) to moderate (0.066). The only cases of between-genome distances (Table 3) ≤ 0.066 are δ(HSV1, HSV2) = 0.023, δ(HHV6, EHV1) = 0.059, δ(HHV6, HSV1) = 0.063, δ(HHV6, HSV2) = 0.064, and δ(HHV6, MDV) = 0.065. These data suggest that HHV6 is quite central among α- and β-HV sequences. The foregoing analysis shows that individual herpesvirus genomes tend to be more homogeneous compared with the between-herpesvirus distances.

In comparing the U_L or U_S sequences between viruses (Table 5), the distances between U_S regions are almost always larger than the distances between U_L regions. Moreover, distances between a U_L region and a U_S region from different viruses are relatively large. This finding reflects the conserved core of herpesvirus common genes in the U_L regions of α-HV sequences.

Partial orderings with respect to standards. The genomic distance evaluations [$\delta(f,g)$ of equation 1] are based on a single number, the absolute average difference of relative abundances $|\rho^*_{XY}(f) - \rho^*_{XY}(g)|$ of the two sequences being compared. To avoid the possibility of a few extreme dinucleotide relative abundances exerting a large influence on the value of $\delta(f,g)$, we introduce a method of partial orderings. In this context, each sequence is represented by the vector of its 16 dinucleotide relative abundances (ρ^*_{XY}). The dinucleotide relative abundance vectors of two genomes are compared with a corresponding 16-component vector of a sequence standard S. If one of the two genomes A and B, say A, is closer to the standard S in at least 13 of the 16 components, a dominance

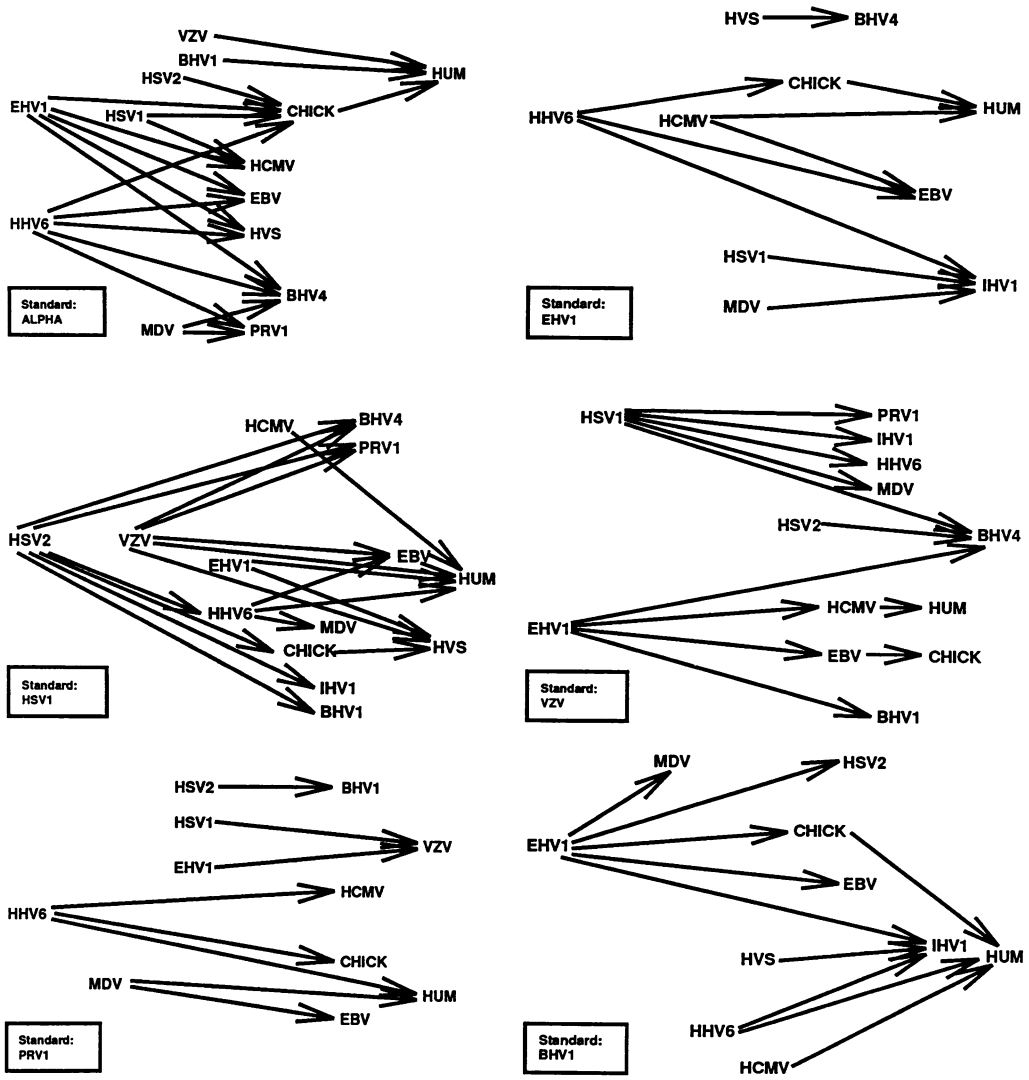

FIG. 2—*Continued.*

ordering between the two genomes relative to the standard is determined, expressed as A dominates B. These evaluations relative to the standard provide a partial ordering based on degree of similarity in dinucleotide relative abundances among the sequences. For a given standard, the closest sequences are those which are undominated and dominate several other sequences; the most distant sequences are those that are dominated by several sequences but dominate none (53).

We use several classes of standards: (i) each individual herpesvirus sequence; (ii) consensus sequences assembled from different groups of herpesvirus sequences (see Materials and Methods); (iii) two host sequence sets, human and chicken; and (iv) an absolute sequence (the unit sequence). With each standard, the comparisons are made for every pair of sequences (Fig. 2). Some observations on these figures are highlighted below.

(i) Human standard (Fig. 2). Not surprisingly, the other vertebrate sequence, that of chickens, is closest to the human sequence and dominates all viral sequences. Next closest is EBV, which dominates 10 sequences, including the BHV4 sequence and all α-HV sequences.

(ii) Chicken standard (data not shown). The undominated human sequence is closest to the chicken sequence, and again EBV dominates most other sequences.

(iii) γ-HV consensus standard (Fig. 2). As expected, the individual γ-HV sequences are closest to this standard. EBV and BHV4 dominate many sequences, whereas HVS dominates only IHV1. Thus, EBV and BHV4 are closer to the γ-HV consensus standard and to each other than either is to HVS.

(iv) EBV standard (Fig. 2). The closest are the human and chicken sequences. Chicken (but not human) dominates HVS. Interestingly, MDV is undominated, suggesting that MDV has some attributes of a γ-HV sequence. This relationship is also consistent with its position to the γ-HV standard, where it fell in the middle level. HHV6, EHV1, and HVS belong to the second tier. The most dominated are the α-HV sequences.

(v) HVS standard (data not shown). EBV is not comparable to any other sequence, although both are classified as γ-HV sequences, reflecting that EBV has a mixed relationship to HVS, being relatively close in certain dinucleotide relative abundances (e.g., CG, TA, and TG/CA) but relatively far with respect to most others.

TABLE 6. Kendall tau correlations (multiplied by 100) of distance orders of 20 standards[a]

	Kendall tau correlation (%)																		
	Human	Chicken	Gamma	EBV	BHV4	HVS	Beta	HCMV	HHV6	Alpha	EHV1	HSV1	HSV2	VZV	MDV	PRV1	BHV1	IHV1	Virus	Unit
Human	100	81	77	83	85	60	−17	−19	−14	−27	−8	−30	−22	−30	−8	−23	−25	−28	−19	3
Chicken	81	100	77	77	75	66	−13	−10	−8	−14	1	−20	−17	−18	−2	−20	−9	−34	−13	7
Gamma	77	77	100	74	74	56	−17	−15	−14	−11	−1	−22	−17	−16	0	−31	−12	−43	−4	−1
EBV	83	77	74	100	84	61	−10	−8	−2	−10	−1	−17	−12	−14	3	−15	−13	−20	−2	9
BHV4	85	75	74	84	100	53	−26	−24	−14	−25	−12	−31	−23	−25	−8	−23	−28	−28	−17	−1
HVS	60	66	56	61	53	100	−17	−13	−25	−37	−17	−45	−46	−42	−29	−33	−7	−56	−34	−15
Beta	−17	−13	−17	−10	−26	−17	100	90	84	70	76	73	72	68	69	61	74	51	71	75
HCMV	−19	−10	−15	−8	−24	−13	90	100	81	69	76	70	70	63	68	67	73	49	66	76
HHV6	−14	−8	−14	−2	−14	−25	84	81	100	73	69	77	77	73	70	72	64	62	72	81
Alpha	−27	−14	−11	−10	−25	−37	70	69	73	100	73	88	87	92	76	51	63	56	86	66
EHV1	−8	1	−1	−1	−12	−17	76	76	69	73	100	70	70	68	79	60	81	49	73	79
HSV1	−30	−20	−22	−17	−31	−45	73	70	77	88	70	100	93	88	81	58	66	57	77	63
HSV2	−22	−17	−17	−12	−23	−46	72	70	77	87	70	93	100	87	78	57	58	65	81	68
VZV	−30	−18	−16	−14	−25	−42	68	63	73	92	68	88	87	100	73	46	62	52	86	63
MDV	−8	−2	0	3	−8	−29	69	68	70	76	79	81	78	73	100	58	68	55	75	73
PRV1	−23	−20	−31	−15	−23	−33	61	67	72	51	60	58	57	46	58	100	39	73	47	65
BHV1	−25	−9	−12	−13	−28	−7	74	73	64	63	81	66	58	62	68	39	100	24	65	64
IHV1	−28	−34	−43	−20	−28	−56	51	49	62	56	49	57	65	52	55	73	24	100	52	59
Virus	−19	−13	−4	−2	−17	−34	71	66	72	86	73	77	81	86	75	47	65	52	100	73
Unit	3	7	−1	9	−1	−15	75	76	81	66	79	63	68	63	73	65	64	59	73	100

[a] See Materials and Methods for complete explanations.

(vi) BHV4 standard (Fig. 2). EBV dominates most other sequences, including human and chicken sequences, and therefore can be regarded as closest to BHV4. HVS is not comparable. This anomaly is consistent with the subdivision of the gamma class into γ_1 and γ_2 (80), with EBV and BHV4 representative of the γ_1 subclass and HVS typical of the γ_2 subclass. Also consistent with other data presented above, the second level entails MDV and EHV1 (both undominated), HHV6, and HSV1. Farthest in the partial orderings is BHV1.

(vii) Consensus β-HV standard (Fig. 2). About equally close are HHV6, HCMV, and EHV1, with many dominance relations, HHV6 having the most.

(viii) HHV6 standard (Fig. 2). This standard in many respects is central among all α- and β-HV sequences. Close to this standard with several dominance orderings are HCMV, EHV1, HSV1, and HSV2. The most dominated sequences include all γ-HV and, in addition, IHV1 and BHV1.

(ix) HCMV standard (data not shown). About equally close are HHV6 and EHV1.

(x) Consensus α-HV standard (Fig. 2). The closest are EHV1 and HHV6. The most distant are the γ-HV sequences, with the human sequence being the most dominated. The chicken sequence dominates the human sequence.

(xi) EHV1 standard (Fig. 2). The closest undominated sequences are those of HHV6 and HCMV, with HHV6 dominating more than HCMV. The most distant sequences are those from the gamma class.

(xii) HSV1 standard (Fig. 2). The closest sequence is HSV2. Other undominated genomes include the alpha types VZV and EHV1. The most dominated are all γ-HV sequences. PRV1 in the α-HV class is substantially dominated.

(xiii) HSV2 standard (data not shown). The closest is HSV1, which also dominates VZV, suggesting that VZV is closer to HSV1 than to HSV2.

(xiv) VZV standard (Fig. 2). The closest are HSV1 and EHV1. Note that HSV1 dominates five sequences. HSV2 is undominated but dominates only a single other sequence, supporting the proposition that VZV is closer to HSV1 than to HSV2.

(xv) MDV standard (data not shown). The undominated group includes HHV6, EHV1, and HSV1. The dominated sequences consist mostly of outlier herpesvirus sequences, the most extreme being HVS and BHV1.

(xvi) PRV1 standard (Fig. 2). The closest to PRV1 is HHV6, and next is MDV. The farthest are several members of the gamma class and curiously VZV.

(xvii) BHV1 standard (Fig. 2). The closest is EHV1, which exhibits many dominance orderings.

Concordance correlations of the distance orderings between genomes. Table 6 reports the $\tau(s,t)$ correlation values of the δ distance orderings (see Materials and Methods) for all pairs among the 20 different sequences.

(i) The γ-HV versus non-γ-HV partition. The τ values partition the distance orderings into two main groups, γ-HV versus non-γ-HV. Between members of different groups, the τ values are essentially zero or slightly negative, while within the groups, the τ values range from about 0.4 up to 0.9. This reflects the fact that among herpesviruses the γ-HV form a separate group, with each member quite distant to any α- or β-HV sequence.

(ii) Within the γ-HV. The high τ value of 0.84 for τ(BHV4, EBV) contrasts with the values of 0.53 and 0.61 for τ(HVS, BHV4) and τ(HVS, EBV), respectively. In agreement with other means of comparing, this again argues for greater similarity between BHV4 and EBV as opposed to BHV4 and HVS. The very high τ concordance correlation (0.85) of BHV4 with the human sequence is intriguing.

(iii) Within the non-γ-HV. An almost perfect τ correlation occurs between HSV1 and HSV2, closely followed by the high τ correlation of VZV with each of them. The core α-HV sequence, EHV1, carries a moderate τ value of about 0.70 to these three. MDV also shows moderate τ values, 0.7 to 0.8, with the four central α-HV members. The two peripheral α-HV BHV1 and PRV1 (having the least overall DNA similarity to other α- and β-HV sequences of our collection) show the lowest τ values with each other in the α-HV group, τ only 0.39. The two β-HV HCMV and HHV6 are highly correlated to one another and moderately correlated (0.6 to 0.8) to the

TABLE 7. Protein sequences analyzed by amino acid scoring alignments

Virus	Homologous gene, corresponding protein length (amino acids)[a]			
	DPOL	VCAP	VTER	VGLB
EBV	BALF5, 1,015	BCLF1, 1,381	BGRF1 + BDRF1, 690	BALF4, 857
HVS	9[b]/KCRF2, 1,009	25, 1,371	29, 683	8/KCRF1, 808
HCMV	UL54, 1,242	UL86, 1,370	UL89, 674	UL55, 906
HHV6	1,012	ORF4L, 1,345	ORF 12L, 667	Fragment, 259
EHV1	30, 1,220	42, 1,376	44, 734	GB/GP14/33, 980
HSV1	UL30, 1,235	UL19, 1,374	UL15, 735	UL27, 904
VZV	28, 1,194	40, 1,396	45 + 42, 747	31, 868
BHV1				GI, 928
MDV				GB, 865
PRV1		MCP142, 1,330		GII, 913
IHV1	57, 985		62 + 69 + 71, 852	

[a] DPOL, DNA polymerase; VCAP, major capsid protein; VTER, probable DNA-packaging (tegument) protein; VGLB, gB precursor; +, spliced. All sequences are in the SwissProt data bank (version 25).
[b] Gene number.

α-HV sequences, including BHV1 and PRV1. The unclassified IHV1 is weakly or not correlated to any of the others, having its highest τ value with PRV1.

Protein sequence comparisons. We applied the PAM120, BLOSUM 56, and SISS scoring protocols (see Materials and Methods), comparing four protein sequences common to the genomes of herpesviruses (Table 7). The resulting SAS scores are given in Table 8. The conclusions were qualitatively the same for the three programs; therefore, only the SAS scores of the PAM120 output are reported.

(i) DNA polymerase. The length of the protein is about 1,010 residues in the γ-HV, compared with about 1,200 residues in the average α-HV. The length in HHV6 is like that of a γ-HV, and the length in HCMV like that of an α-HV. The within-α-HV class, SAS scores (see Materials and Methods) range from 2.37 to 2.90, of the same magnitude as that of the

TABLE 8. Protein sequence comparisons among herpesviruses

Protein	Virus	Length (bp)	PAM120 SAS score[a]								
			EBV	HVS	HCMV	HHV6	EHV1	HSV1	VZV	PRV1	MDV
DNA polymerase	EBV	1,015									
	HVS	1,009	2.85								
	HCMV	1,242	1.71	1.65							
	HHV6	1,012	1.30	1.58	2.28						
	EHV1	1,220	1.60	1.79	1.14	1.23					
	HSV1	1,235	1.62	1.62	1.04	1.20	2.90				
	VZV	1,194	1.53	1.63	1.15	1.38	2.68	2.37			
Major capsid	EBV	1,381									
	HVS	1,371	3.09								
	HCMV	1,370	1.00	1.19							
	HHV6	1,345	1.05	0.84	2.47						
	EHV1	1,376	0.93	0.90	0.54	0.78					
	HSV1	1,374	0.85	0.83	0.54	0.68	2.90				
	VZV	1,396	0.84	0.86	0.55	0.74	3.31	2.78			
	PRV1	1,330	0.90	0.86	0.62	0.69	3.98	3.08	3.22		
Tegument	EBV	690									
	HVS	683	2.89								
	HCMV	674	1.66	1.51							
	HHV6	667	1.55	1.28	2.95						
	EHV1	734	1.53	1.58	1.65	1.52					
	HSV1	735	1.55	1.47	1.74	1.75	3.44				
	VZV	747	1.32	1.31	1.57	1.60	3.29	3.25			
gB	EBV	857									
	HVS	808	2.00								
	HCMV	906	1.13	1.09							
	HHV6	259	1.22	1.06	1.48						
	EHV1	980	0.81	0.89	0.77	0.86					
	HSV1	904	0.93	0.82	0.87	0.70	2.52				
	VZV	868	0.88	0.80	0.85	0.65	2.76	2.53			
	PRV1	913	0.88	0.68	0.78	0.79	2.83	2.62	2.89		
	MDV	865	0.80	0.88	0.90	0.69	2.28	2.54	2.54	2.20	
	BHV1	928	0.69	0.72	0.68	0.80	2.77	2.39	2.78	2.84	2.36

[a] See Materials and Methods for details of the scoring protocol.

TABLE 9. Comparisons of herpesviruses with adenovirus and vaccinia virus for the DNA polymerase gene

Virus	Length (bp)	SAS score[a]	
		Adenovirus (1,056 bp)	Vaccinia virus (1,006 bp)
EBV	1,015	0.06	0.19
HVS	1,009	0.05	0.23
HCMV	1,242	0.00	0.25
HHV6	1,012	0.00	0.28
EHV1	1,220	0.00	0.23
HSV1	1,235	0.05	0.21
VZV	1,240	0.00	0.31
Vaccinia virus	1,006	0.00	

[a] See Materials and Methods for details.

within-γ-HV class (EBV versus HVS; SAS score = 2.85). The DNA polymerase SAS score is highest between EHV1 compared with HSV1, about the same as the EBV-versus-HVS SAS score. For the within-β-HV class (HCMV versus HHV6), the SAS score is unambiguously the smallest (2.28) compared with the within-γ-HV class and within-α-HV class SAS scores. The SAS scores for DNA polymerase between a γ-HV sequence and each corresponding α- or β-HV sequence are mostly in the range 1.30 to 1.79, at a level about 50% reduced from the within-α-HV class or within-γ-HV class SAS scores. The SAS scores for each β-HV sequence compared with an α-HV sequence are smaller, about 1.11 for HCMV and 1.27 for HHV6. Thus, the DNA polymerase protein within-class comparisons entail SAS scores 50 to 100% higher than the between-class comparisons.

Earl et al. (27) used the FASTA alignment program to assess amino acid similarity of the DNA polymerase sequences among HSV1, adenovirus, and vaccinia virus and claim some significant similarity (see also reference 86). The SAS score of the DNA polymerase gene of herpesviruses versus adenovirus are not significant (Table 9). The comparisons of the DNA polymerase sequence between herpesvirus and vaccinia virus achieve a SAS score less than one-third of the smallest SAS score among the herpesvirus sequences. These comparisons with vaccinia virus obviously do not discriminate among α-, β-, or γ-HV classes, indicating pronounced deviation of vaccinia virus from all herpesviruses.

(ii) Major capsid protein. The major capsid gene is also available for PRV1. The length is approximately constant (1,370 residues) for all viruses. The within-γ-HV class SAS score (3.09) was in reasonable agreement with the corresponding SAS scores (2.78 to 3.98) among α-HV sequences. The smallest SAS score was for HSV1 versus VZV (2.78). Surprisingly high was the SAS score of 3.98 attained for the major capsid proteins of EHV1 versus PRV1.

The within-β-HV class SAS score (2.47) was significantly less than that of the within-γ- and α-HV class SAS scores. The SAS scores of γ-HV sequences against β- or α-HV sequences were in the range of 0.83 to 1.05. The SAS score for a β-HV versus α-HV sequence was smaller, in the range of 0.54 to 0.78. Thus, the between-class SAS scores were about 30 to 50% smaller than the within-class SAS scores.

(iii) Tegument protein. The length of the tegument protein is almost the same for the β- and γ-HV sequences (about 680 residues) and about 50 residues longer in the α-HV class (Table 7). The within-class SAS scores of β- and γ-HV sequences are essentially the same (about 2.9), and the within-α-HV class analysis yields the higher score range, 3.25 to 3.44. The between-class SAS scores of γ-HV versus α- and β-HV sequences and β-HV versus α-HV sequences are about 50% of the within-class SAS scores.

(iv) Glycoprotein gB. The gB sequence was available from 10 herpesvirus genomes. The length of the gene averages about 900 residues, slightly less in HVS, and only a fragment of length 259 residues was available from HHV6. Among the α-HV sequences, the SAS scores were in good agreement (2.20 to 2.89). The within-γ-HV class SAS score was 2.00, while the between-γ- and α-HV sequence comparisons yielded scores in the range of 0.68 to 0.93. The comparison of γ-HV sequences with the HCMV and HHV6 representatives gave higher SAS scores, around 1.10. The SAS score for the β-HV sequences versus each α-HV sequence yielded SAS scores in the range of 0.65 to 0.90, about the same as for a γ-HV sequence compared with an α-HV sequence.

DISCUSSION

Herpesviruses are widespread in vertebrate species, sharing several moderately to well conserved genes, as determined from amino acid identity comparisons (e.g., DNA polymerase and gB) even though they exhibit a dramatic variation in mean G+C genomic frequency ranging from 35 to 75% (43). The herpesviruses are believed to be of ancient origin, at least 300 million years old (23). They have coevolved with their hosts, mainly mammals, birds, and fish, with possible multiphyletic origins and lateral transfers.

Most studies of evolutionary development and of phylogenetic (tree) reconstructions among groups of organisms are based on protein sequence comparisons. The caveat has been that different trees may result for the same set of organisms based on analysis of different proteins and dependent on the alignment algorithm (58, 68). The practice of phylogenetic reconstruction recounts many examples where different sets of proteins produce different phylogenies. This is the case for phylogenies constructed around structural proteins, immunological typings, and comparisons of local genomic regions or rRNA genes. Phylogenies constructed from individual genes are interesting insofar as the evolution of these specific genes is concerned, but their extrapolation to the evolutionary relationship between, say, viruses probably requires analysis of complete genomes, not just individual proteins. In viruses, a further caveat applies: the gene in question may have undergone direct transfer between genomes during its evolution. Moreover, the genome in question may be a genetic mosaic, having acquired genes from different sources and undergone lateral transfer, transposition, and recombination events in the course of evolution. Phylogenies determined by a single type of gene have to be evaluated with cognizance of all other available data.

We have introduced novel distance measures and methods (for rationale, see below) based on dinucleotide relative abundances of genomic sequences from which to derive evolutionary relationships among herpesvirus genomes. For corresponding studies comparing simultaneously di-, tri-, and tetranucleotide relative abundance distances, see reference 52. These methods of oligonucleotide relative abundance distances have been applied in comparing complete genomes of coliphage (9) and in extensive analyses of sequences from more than 40 diverse prokaryotic and eukaryotic species (52, 55). In the second part of this study, recent methods of protein sequence alignments are implemented in comparing four protein sequences among herpesviruses (Table 8).

Genomic sequences were compared in several ways: (i) with respect to similarities and differences of genomic composi-

tional extremes (Table 2); (ii) with respect to distance orderings within and between herpesvirus genomes (Tables 3 to 5; Fig. 1); (iii) with respect to partial orderings for the vector of dinucleotide relative abundances of two herpesvirus genomes to appropriate sequence standards (Fig. 2); (iv) with respect to similarity of an individual herpesvirus sequence to a consensus α-, β-, or γ-HV sequence; and (v) with respect to concordances and discordances of correlations of the distance orderings for pairs of HV sequences (Kendall tau correlations; Table 6).

We highlight below some of the main results emanating from our genomic comparisons and from our protein sequence comparisons. There are invariants and contrasts for which we will venture some interpretations.

Centrality of the HHV6 and EHV1 genomes. The sequence with the most dominances in the partial orderings (Fig. 2) cumulated over all standards is HHV6, with the next in this dominance hierarchy being EHV1. These viruses lie closest to the α-HV consensus standard and about equally close to the β-HV consensus sequence. The global consensus herpesvirus standard also has HHV6 and EHV1 significantly close. HHV6 is also close to EBV with respect to many standards (e.g., human, HSV2, EHV1, and consensus α-HV). These results argue for the centrality of the HHV6 genome. Along these lines, it is known that HHV6 is lymphotropic and can infect T cells but it is also associated with monocytes, epithelial cells, and the central nervous system (36, 54a, 54b, 78a). On the basis of the protein comparisons, we propose that essentially all α-HV are of more recent origin than the β- and γ-HV and that HHV6 might be akin to a progenitor herpesvirus.

Relative CpG deficiencies in herpesvirus genomes. All γ-HV have the dinucleotide CpG significantly underabundant and concomitantly carry a relative excess of CpA/TpG (Table 2) both to the same extent as in vertebrate genomes. This characteristic is not exhibited by α- or β-HV. Honess et al. (45) proposed that the standard methylase/deamination/mutation hypothesis should apply to viruses found in highly dividing cells. However, MDV and HHV6, which present some biological features of γ-HV (lymphotropism), do not show CpG suppression nor any biased dinucleotide relative abundance (Table 2). Paradoxically, CpG is high only in HCMV dinucleotide relative abundance over its whole genome (Table 2). Apropos, all vertebrate and invertebrate mitochondrial genomes of metazoan species are strongly CpG suppressed (data not shown). In the latter, it seems unlikely that methylation would be involved in this suppression, since invertebrates (e.g., *D. melanogaster*, *C. elegans*, and sea urchins) apparently do not possess any known relevant methylase. There appear to be other mechanisms or factors which lead to CpG depletion (52).

Closeness between EBV and human sequences. With respect to the human sequence standard in the partial orderings, EBV dominates almost all other herpesviruses, including all α-HV sequences, HHV6, and BHV4. Moreover, distance ordering arrays relative to the human and EBV standards are highly concordant according to the Kendall τ correlation coefficient (Table 6). These observations are consistent with the possible acquisition of about one-third of the EBV genome, including all of the latent genes from a host cell (in this connection, see the detailed analysis of codon preferences in EBV genes in reference 50; see also reference 30).

Similarities of herpesvirus sequences to avian versus mammalian host sequences. For every herpesvirus standard, the chicken sequence dominates all sequences that the human sequence dominates. Moreover, there are many standards for which the chicken sequence dominates the human sequence, e.g., the EHV1, HCMV, consensus herpesvirus, and consensus α-HV standards. In addition, the chicken (but never the human) sequence dominates HVS for several of the herpesvirus standards. Human never dominates any herpesvirus sequence for an α- or β-HV sequence standard. From these relationships, it is tempting to speculate that many herpesvirus antecedents inhabited avian species and later underwent transfer from birds into mammals. There is precedent for such an evolutionary scenario in the case of the influenza A virus transmutations (33, 35).

Classification of BHV4. Originally, BHV4 was assigned to the β-HV class, but a closer association with the γ-HV class was shown when sequence data became available (12, 13, 83). From their analysis in comparing a number of specific genes, Ehlers et al. (29) and Bublot et al. (12, 13) suggest that BHV4 is evolutionarily closer to HVS than to EBV. Our DNA comparisons based on the distance orderings in Table 3 and especially the partial orderings (Fig. 2) and also the Kendall tau correlations (Table 6) all point to the opposite relationship, indicating that BHV4 is much closer to EBV than to HVS.

Very distant herpesvirus genomes. Although the two bovine herpesviruses infect the same host, BHV1 and BHV4 are separated by the greatest distance among the examples of Table 3. One possible explanation involves the molecular resource partitioning principle (see below). Aside from IHV1, HVS and BHV1 tend to be more dominated than any other herpesvirus sequence in the partial orderings.

Is VZV closer to HSV1 or to HSV2? The partial orderings persistently attest to greater similarity of VZV to HSV1 than to HSV2. Thus, closest to the VZV standard are HSV1 (which dominates HSV2) and EHV1. Also, the partial ordering based on the HSV2 standard has VZV dominated by HSV1. VZV tends to be dominated with respect to all β-HV sequence standards. Telford et al. (81) argue that EHV1 genes are in general more closely related in amino acid sequence to VZV than to their HSV1 counterparts. Our results using DNA sequence distances and partial orderings do not agree with this conclusion.

How is MDV to be classified? The partial orderings for the MDV standard entail the fewest dominance relationships. Technically, this signifies that the dinucleotide relative abundance array of MDV is close to most other HV genomes in several components (at least four) and far in several components (at least four). This may reflect on MDV's unusual combination of a lymphotropic virus with α-HV-like genomic organization. Interestingly, MDV is undominated from the EBV standard, further suggesting that MDV has some attributes of a γ-HV. On the basis of analysis of the protein sequence gD, Ross and Binns (72) argue that MDV is reasonably related to both α- and γ-HV sequences. The phylogenetic tree derived from comparisons of the gD homolog of HSV1 supports the hypothesis that MDV is more closely related to the α-HV family than to the γ-HV family (4). This conclusion may be influenced by the fact that gD is encoded in the U_S region. However, our general DNA partial ordering analysis places MDV closest to HHV6 rather than to any α-HV sequence.

Herpesvirus sequences with the fewest dinucleotide compositional extremes. The most random herpesvirus genome [the sequence of smallest $\delta(1,g)$ distance for all individual and consensus herpesvirus sequences] is EHV1; the second most random herpesvirus genomes are about equally those for MDV and HHV6. In general, the α-HV tend to be more random, i.e., closer to the unit standard, than the members of the γ-HV class. The distance of a herpesvirus genome with a randomly shuffled version of itself would be essentially the same as the distance to the unit standard.

Relatedness of α- and β-HV genomic sequences. In terms of

global DNA sequence distances, the β- and α-HV sequences are relatively close and seem to overlap in some analyses (Fig. 1). The closest distance to the PRV1 genome are the β-HV. The Kendall τ coefficients (Table 6) give further indication that the core α- and β-HV sequences correlate significantly. This conclusion based on genomic comparisons differs from that from our protein sequence comparisons (see below).

γ-HV subclass types 1 and 2. The prototype of type 1 is EBV, and the prototype of type 2 is HVS (80). Members of the γ-HV group (EBV, HVS, and BHV4) are substantially diverse but in aggregate are very distant from the α- and β-HV classes (Fig. 1). Genes involved in important biological properties such as latency, immortalization, gene regulation, and other virus-host interactions are not conserved between HVS and EBV (22), and EBV and HVS have been shown to be more distant than HSV1 is to VZV (1). This is in agreement with our genomic DNA distance assessments δ(EBV, HVS) = 0.155 versus δ(HSV1, VZV) = 0.081. In comparing similarities between specific genes, e.g., the tegument protein and gB genes, the differences persist, while for other key genes, e.g., the DNA polymerase and major capsid protein genes, the similarity scores are about the same (Table 8). The most distant on the amino acid level are the β-HV sequences, although this is not the case when DNA distances are compared.

Similarity among U_S regions of herpesviruses. There is much gene colinearity in direct and/or in inverted order among the U_L sections of the α-HV but less among the U_S sections. Previous reports have concluded that "the close colinear relationship between α-herpesvirus genomes applies strictly only to the L segment" (22). U_S appears to be the more divergent component of α-HV (Table 5). On the basis of comparisons of amino acid sequences in open reading frames, McGeoch (65) suggested that each of the seven distinct genes of the U_S region of VZV has a homolog in HSV1, whereas other genes in the U_S segment of HSV1 do not. However, the similarity of the putative homologs (except for ICP4 and US3) are very weak (74). The most homogeneous genome is HCMV, δ(HCMV U_L, HCMV U_S) = 0.032 (Table 4), substantially smaller than the U_L-versus-U_S distances of EHV1 and of HSV1, $\delta \geq 0.062$. Could this be a consequence of the generation of gene families in HCMV (not present in EHV1 and HSV1) that intermingle the U_L and U_S regions (19, 37)?

Evolution of U_S regions. By what processes did the U_S sequences of α-HV originate and develop? That U_S regions evolved more recently was suggested by Telford et al. (81) since, except for HCMV, similar counterparts are not found in β- or γ-HV sequences. Even with the α-HV, the U_S segments of different viruses, excluding the internal and terminal repeats, bear little significant similarity to each other and no similarity to the U_L sections (except for the gene families of HCMV). We propose that the U_S sections evolved as or resulted from a transposon or retrotransposon insertion or some form of nonhomologous recombination into the herpesvirus genome at or near the terminal repeats during (or perhaps following) circularization in conjunction with viral replication. The terminal repeats and nearby regions tend to be active, often displaying a complex of local sequence iterations and inversions, and might naturally accept transpositions of all kinds. Mobile elements often carry genes between their terminal repeats. Because of the central position of EHV1 (evident from DNA comparisons), we might envision this genome or a close relative as among the progenitor α-HV in which a primordial U_S was acquired and transmitted to other α-HV.

What evidence is there to support the hypothesis that the U_S section arose as a series of rare transposon-like insertions? Genes from transposons (as with genes in plasmids of bacteria) are basically not essential for growth but could encode functions needed on occasion, and some of these genes could affect viral or cellular phenotype (16). There are many cases of genes recently acquired from external sources in herpesvirus genomes, e.g., *DHFR* in HVS (85), the BcRF1 (= interleukin-10 [IL-10]) gene in EBV (45a), gene 13-TS in VZV (83a) and HVS (44), and chemokine receptor, major histocompatibility complex class I-like, and T-cell receptor-like genes in HCMV, putatively captured via mobile transposition aided by recombination events. Recent analysis has revealed several reading frames of HVS with strong sequence similarities to known cellular genes, including G_1-cyclin and IL-8 receptor genes, a T-lymphocyte-activating gene, complement inhibitory factor genes, and a number of cellular U-RNA genes (34). Moreover, there is some evidence that HHV6 may have acquired genes by the same process used by host cell chromosomes to add their termini (60). Also, there is evidence for a transposable insertion in MDV (47). Retrovirus (reticuloendotheliosis virus) insertions into MDV appear to be common and could occur shortly following coinfection of reticuloendotheliosis virus and MDV. These insertions cluster at the IR_S/U_S junction, an active region of the MDV genome, which shows substantial size heterogeneity over several strains (47).

Within- versus between-dinucleotide relative abundance distances. In view of the distances between U_L and U_S generally being smaller than those between herpesvirus genomes (see Results), it is conceivable that by some train of historical events (e.g., homologous and nonhomologous recombination at the terminal repeats or some kind of fusion mechanism), two U_L segments might be combined into two isomeric forms subject to subsequent expansions, contractions, and homogenizations (concerted evolution). This may have occurred several times and most recently over the time frame of herpesvirus evolution in HCMV and VZV, which may account for greater DNA structural (dinucleotide relative abundance) similarity of U_L with U_S in these genomes (Tables 4 and 5). HCMV seems especially vulnerable to complex genomic rearrangements and amplifications. Along these lines, the DNA of HSV1 in infected cells undergoes high levels of homologous recombination linked to DNA synthesis (18). The idea that the HSV1 genome arose by fusion of two genomes (precursors of the L and S components) and that the S component evolved quite differently in each virus family is implicit in the study of Hayward et al. (40) and discussed further by Roizman (70, 71).

Gene acquisitions, losses, and lateral transfer. To various degrees, herpesviruses are replete with complex direct and inverted repeat structures putatively responding to immune system challenges of the host and engaging in transfer of DNA between different hosts and other viruses. It has been proposed (85) that the *DHFR* gene of the lymphotropic herpesvirus HVS is a recent acquisition into the HVS genome not likely to be essential for viral productive growth. The *DHFR* gene of HVS is about 90% identical to the primate version, strongly suggesting that a host gene copy of *DHFR* was recently acquired by HVS. The BcRF1 gene of EBV is significantly similar to the human IL-10 gene. Both the viral IL-10 gene of EBV and the TS gene of VZV are about 70% similar to their human versions. A strong homolog of the TS genes has recently been identified in EHV2 (55% G+C) and EHV5 (53% G+C) of the γ-HV class (80). Since these genomes are somewhat G+C rich, the putative role (44) of TS in VZV and HVS in enhancing A+T-directed substitutions seems doubtful. From this perspective, TS is likely a nonessential acquired gene. For other putative examples of lateral transfer of DNA between different

hosts and other viruses, see references 7 and 61. The foregoing examples attest to the dynamic character of herpesvirus genomes.

Molecular resource partitioning. The exclusion principle of ecological community theory states that coexisting life forms strive to establish independent niche breadth with respect to resources and associated mechanisms (e.g., see text of Roughgarden [73] on the Gause exclusion principle). It is proposed that the otherwise similar α-HV, HSV1 and VZV, have evolved divergent base compositions and patterns of codon usage in part to avoid competition for host resources such as nucleic acid precursors, aminoacyl-tRNAs, and other cellular processes and machinery (74). The avoidance of competition between the ancestral HSV1 and VZV viruses may also have led to spatial and/or temporal isolation. In this connection, we note that the replication of HSV1 has been observed to occur in a few special cellular locations (69) and that the epithelial sites of reactivation and the sites of latency of HSV1 and VZV are distinct (21). The large distance (the largest distance among all herpesviruses analyzed) between BHV1 and BHV4 may also be a phenomenon of molecular resource partitioning. Paraphrasing ecological developments, there may be advantages to divergent molecular niche associations for transcriptional, translational, replication, and genomic segregation objectives together with restrictions on the number of competing species that a molecular environment can sustain. An ancestral herpesvirus may have undergone adaptive radiation in order to occupy the diverse array of molecular niches presented by the host organism.

What are possible mechanisms and events affecting the course of the evolutionary development of herpesviruses? During or following phylogenetic speciation from an ancestral herpesvirus (which we propose to be like HHV6), various selective and historical events evoked changes in life processes and behavior that are also reflected in alterations of DNA and some amino acid sequences. Such events could include horizontal transfer among host species or cell types; altered modalities (such as replication at different times in different tissues, cell types, or compartments of cells) coupled to cellular systems during both the latent and lytic cycles; gain or loss of special genes and control elements; major genomic deletions, transpositions, amplifications, or rearrangements; or novel interactions with host macromolecules. From the perspective of an invading herpesvirus, a human (or other host) cell must appear as a rich and varied habitat, replete with niches compatible with transcriptional, translational, and replicational objectives of the virus. In this context, we note that the human and all vertebrate genomes are quite heterogeneous, with alternating compartments of high and low G+C content (5, 49) and a gene collection with a multimodal codon distribution (50), possibly reflecting gene subclasses which replicate at different times and/or nuclear locations.

What do distances between genome sequences defined in terms of dinucleotide relative abundances discriminate? The distance measure between DNA sequences based on dinucleotide relative abundances (equation 1) rather than straight dinucleotide frequencies (equation 2) seems to work very well and to provide meaningful and revealing comparisons. It is not entirely clear why this should be the case, but it should be noted that many factors that influence DNA oligonucleotide composition and structures have been identified; these include dinucleotide thermodynamic stacking energies (11, 26), DNA packaging, conservation, and variation of codon positions [(I, II), (II, III) or (III, I)], methylation and other nucleotide modifications of various kinds, including methylase of restriction systems in prokaryotes, the standard methylase in vertebrates, and RIPping in *N. crassa* and *Ascobolus* spp. (75). All of these factors could affect dinucleotide relative abundances in different ways and give clues about phylogenetic relationships. In addition, there is some evidence that dinucleotides may be of vital importance in determining local DNA structure (15, 87).

Dinucleotide relative abundances effectively assess residuals of dinucleotide frequencies from those expected from the component mononucleotide frequencies. These deviations may reflect on DNA duplex curvature, twist, roll, and hence supercoiling and other higher-order DNA structural features. DNA structures perhaps more than specific sequences may be essential in modulating processes of replication and repair and in characterizing the salient signature of sequences for these purposes. Many enzymes, especially DNA repair enzymes, recognize lesions or shapes in DNA secondary structures rather than specific sequences. Nucleosome positioning, interactions with DNA-binding proteins, and ribosomal binding of mRNA appear to be strongly affected by dinucleotide arrangements (15, 84, 87). Intimately involved in the generation of the DNA helix are the second-order (15) effects centering on propeller twist and base stacking which are considered driving forces behind many of the relationships between base pairs seen in DNA structure. These authors further emphasize correlations between successive base pairs, especially between roll and slide parameters and cases of bi-stable structures. There are data suggesting that the dinucleotide relative abundance distance reflects largely on commonalities and differences in DNA structures.

DNA has at least two functions: (i) to effect genome replication and segregation and (ii) to encode appropriate gene products. The first may be mostly DNA structure specific, while the second certainly emphasizes sequence specificity. The genome presumably requires flexibility and balance to accomplish both. Evolution of a genome may be constricted by the protein functions it encodes. There are indications that chromosomal replication and segregation depend on DNA stacking in a decondensed structure. In this context, DNA produces loops, which have to be resolved. The replication cycle is not well understood in eukaryote organisms, and the controls may be more structure specific than sequence specific (3). Along these lines, the ρ^* relative abundance values appear to discriminate more structure specificity.

Our discussion continues with interpretations of the protein sequence comparisons.

Protein sequence scores within and between α-, β-, and γ-HV classes. For two of the proteins (gB and tegument), the SAS scores among the α-HV class sequences tend to be about 30 to 100% higher than the within-γ-HV class and within-β-HV class SAS scores. However, for the larger essential proteins, i.e., DNA polymerase and the major capsid protein, the within-α-HV class and within γ-HV class SAS scores have similar magnitudes. The comparisons between α-, β-, and γ-HV sequences with respect to each of the proteins tend to produce SAS scores 30 to 100% lower than the within-class comparisons. The foregoing results support the coherence of the α-, β-, and γ-HV classes and the clear demarcation between classes. The β-HV class (HCMV and HHV6) appears on the protein level to be the least conserved.

Reduced within-β-HV class scores. For all protein types, the within-β-HV class (HCMV versus HHV6) comparisons yield SAS scores significantly lower than those for the within-α-HV class comparisons. The within-β-HV class and within-γ-HV class SAS scores for the tegument protein are in accord. On the basis of our dinucleotide relative abundance distances analysis, we found that the α- and β-HV genomes are quite close;

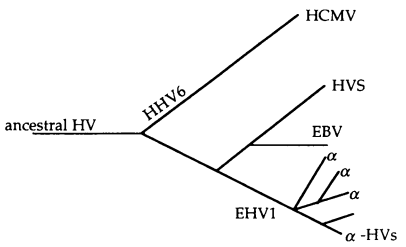

FIG. 3. Protein similarities depicted by tree topology.

especially, the HHV6 and EHV1 genomes are significantly closer than the α- versus γ-HV genomes. This is in contrast to the results of the four protein similarity evaluations which mostly place the β- and γ-HV sequences equally far from the α-HV sequences. Corresponding conclusions are valid (data not shown) for comparisons of the helicase subunit (homolog of UL5 of HSV1) for ribonucleotide reductase (large subunit), for dUTPase, and for thymidine kinase, among others.

Genomic relationships and specific gene comparisons. The DNA polymerase and major capsid proteins presumably change more slowly (are more conserved) relative to gB and the tegument protein, which could account for the almost congruent SAS score of EBV versus HVS with HSV1 versus VZV. Consistent with this interpretation is the observation that the SAS score for the major capsid protein is greater between γ- and β-HV sequences than between γ- and α-HV sequences.

The arrangement of genes flanking the DNA polymerase gene discriminate the α-, β- and γ-HV classes, with the most rearrangements observed in the γ-HV sequences (82). However, we find the SAS score within the γ-HV (EBV versus HVS) to be of about the same magnitude as that among the α-HV sequences (EHV1, HSV1, and VZV) but significantly greater than the score within the β-HV (HHV6 versus HCMV).

The major capsid gene of HSV1 and PRV1 is part of a colinear set of four open reading frames (54). Klupp et al. find that the degree of similarity of this gene is about the same between PRV1 and VZV as between PRV1 and HSV1 (54). In agreement, our evaluation (Table 8) indicates significant similarity, but the strongest similarity of the major capsid protein gene is observed between PRV1 and EHV1, with an intermediate level of similarity for PRV1 with VZV and a further reduced level of similarity for PRV1 with HSV1.

Chou and Marousek (20) observe that the gB protein sequences of HHV6 and HCMV are more similar to each other than to the gB sequences of other herpesviruses. Our analysis of four essential proteins (see Results) suggests that this comparison prevails for all conserved genes, connoting that the within-β-HV group (HHV6 versus HCMV) gene similarity is significantly greater than the average similarity scores between β-HV and corresponding α- or γ-HV sequences.

Some speculations and hypotheses. The diminished SAS scores for the within-β-HV class protein comparisons relative to the within-α- or within-γ-HV class scores and the pervasive lower similarity scores between α-, β-, and γ-HV suggests that HHV6 and HCMV separated earlier than did the other herpesviruses. A pictorial representation of the protein similarities is furnished by the tree topology shown in Fig. 3.

The higher similarity scores in protein sequence comparisons and the closer DNA distances among α-HV supports the hypotheses that α-HV are of more recent ancestry. EHV1 stands out as the most central α-HV, with stronger similarities on both the DNA and protein levels to the other α-HV.

It is interesting to observe that for protein gB all the mutual SAS scores among EHV1, PRV1, BHV1, and VZV exceed 2.75 (Table 8), and the SAS score of the pair HSV1 and BHV2 is 2.82, while all SAS scores between these two groups range from 2.39 to 2.62, projecting a division of the α-HV genomes into two subgroups (67a). The SAS score for the MDV sequence with all other α-HV sequences is generally the lowest.

ACKNOWLEDGMENTS

This work was supported in part by NIH grants 5R01GM10452-30 and 8R01HG00335-05 and NSF grant DMS-9106974 to S.K., NIH grant R01AI20211 to E.S.M., and SAAF grant to G.A.S.

We are happy to acknowledge helpful comments on the manuscript by B. E. Blaisdell, V. Brendel, and A. M. Campbell.

REFERENCES

1. **Albrecht, J. C., and B. Fleckenstein.** 1990. Structural organization of the conserved gene block of herpesvirus saimiri coding for DNA polymerase, glycoprotein B, and major DNA binding protein. Virology **174:**533–542.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. L. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Bembow, R. M., J. Zhao, and D. D. Larson.** 1992. On the nature of origins of DNA replications in eukaryotes. Bioessays **14:**661–670.
4. **Bennett, A. M., L. Harrington, and D. C. Kelly.** 1992. Nucleotide sequence analysis of genes encoding glycoproteins D and J in simian herpes B virus. J. Gen. Virol. **73:**2963–2967.
5. **Bernardi, G., and G. Bernardi.** 1986. Compositional constraints and genome evolution. J. Mol. Evol. **24:**1–11.
6. **Bird, A. P.** 1986. CpG-rich islands and the functions of DNA methylation. Nature (London) **321:**209–213.
7. **Birkenbach, M., K. Josefsen, R. Yolamanchili, G. E. Lenoir, and E. Kieff.** 1993. Epstein-Barr virus-induced genes: first lymphocyte-specific G protein-coupled peptide receptors. J. Virol. **67:**2209–2220.
8. **Bishop, S. E., P. Fienberg, and W. Holland.** 1975. Discrete multivariate analysis: theory and practice. M.I.T. Press, Cambridge, Mass.
9. **Blaisdell, B. E., A. M. Campbell, and S. Karlin.** Unpublished observations.
10. **Brendel, V., I. Ladunga, and S. Karlin.** Significant segment score (SISS) matrices for comparing protein sequences. Submitted for publication.
11. **Breslauer, K. J., R. Frank, H. Blöcker, and L. A. Marky.** 1986. Predicting DNA duplex stability from the base sequence. Proc. Natl. Acad. Sci. USA **83:**3746–3750.
12. **Bublot, M., P. Lomonte, A.-S. LeQuarre, J.-C. Albrecht, J. Nicholas, B. Fleckenstein, P.-P. Pastoret, and E. Thiry.** 1992. Genetic relationships between bovine herpesvirus 4 and the gammaherpesviruses Epstein-Barr virus and herpesvirus saimiri. Virology **190:** 654–665.
13. **Bublot, M., M.-F. Van Bressem, E. Thiry, J. Dubuisson, and P.-P. Pastoret.** 1990. Bovine herpesvirus 4 genome: cloning, mapping and strain variation analysis. J. Gen. Virol. **71:**133–142.
14. **Burge, C., A. M. Campbell, and S. Karlin.** 1992. Over- and under-representation of short oligonucleotides in DNA sequences. Proc. Natl. Acad. Sci. USA **89:**1358–1362.
15. **Calladine, C. R., and H. R. Drew.** 1992. Understanding DNA. Academic Press, San Diego, Calif.
16. **Campbell, A. M.** Genome organization in procaryotes. Curr. Opin. Genet. Dev., in press.
17. **Cavalli-Sforza, L. L., and A. W. F. Edwards.** 1967. Phylogenetic analysis. Models and estimation procedures. Am. J. Hum. Genet. **19:**233–257.
18. **Challberg, M. D., and T. J. Kelly.** 1989. Animal virus DNA replication. Annu. Rev. Biochem. **58:**671–717.
19. **Chee, M. S., A. T. Bankier, S. Beck, R. Bohri, C. M. Brown, R.**

Cemy, T. Horsnell, C. A. Hutchison III, T. Kouzarides, J. A. Martignetti, E. Preddie, S. C. Satchwell, P. Tomlinson, K. M. Weston, and B. G. Barrell. 1990. Analysis of the protein coding content of human cytomegalovirus strain AD169. Curr. Top. Microbiol. Immunol. **154:**125–169.

20. **Chou, S., and G. I. Marousek.** 1992. Homology of the envelope glycoprotein B of human herpesvirus-6 and cytomegalovirus. Virology **191:**523–528.
21. **Croen, K. D., J. M. Ostrove, L. J. Dragovic, and S. E. Straus.** 1988. Patterns of gene expression and sites of latency in human nerve ganglia are different for varicella-zoster and herpes simplex viruses. Proc. Natl. Acad. Sci. USA **85:**9773–9777.
22. **Cullinane, A. A., F. J. Rixon, and A. J. Davison.** 1988. Characterization of the genome of equine herpesvirus 1 subtype 2. J. Gen. Virol. **69:**1575–1590.
23. **Davison, A. J.** 1993. Origins of the herpesviruses, p. S-27. Abstr. Int. Herpesvirus Meet.
24. **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.
25. **Dayhoff, M. O., R. M. Schwartz, and B. C. Orcutt.** 1978. A model of evolutionary change in proteins, p. 345–358. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5, suppl. 3. National Biomedical Research Foundation, Washington, D.C.
26. **Delcourt, S. G., and R. D. Blake.** 1991. Stacking energies in DNA. J. Biol. Chem. **266:**15160–15169.
27. **Earl, P. L., E. V. Jones, and B. Moss.** 1986. Homology between DNA polymerases of poxviruses, herpesviruses, and adenoviruses: nucleotide sequence of the vaccinia virus DNA polymerase gene. Proc. Natl. Acad. Sci. USA **83:**3659–3663.
28. **Eck, R. V., and M. O. Dayhoff.** 1966. Atlas of protein sequence and structure. National Biomedical Research Foundation, Silver Spring, Md.
29. **Ehlers, B., H.-J. Buhk, and H. Ludwig.** 1986. Analysis of bovine cytomegalovirus genome structure: cloning and mapping of the monomeric polyrepetitive DNA unit, and comparison of European and American strains. J. Gen. Virol. **66:**55–68.
30. **Farrell, P. J.** 1993. Epstein-Barr virus as an example of the evolution of the herpesvirus strategy, p. S-24. Abstr. Int. Herpesvirus Meet.
31. **Felsenstein, J.** 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17:**368–373.
32. **Felsenstein, J.** 1988. Phylogenics from molecular sequences: inference and reliability. Annu. Rev. Genet. **22:**521–565.
33. **Fitch, W. M., J. M. E. Leiter, X. Li, and P. Palese.** 1991. Positive Darwinian evolution in human influenza A viruses. Proc. Natl. Acad. Sci. USA **88:**4270–4274.
34. **Fleckenstein, B., J. Albrecht, I. Müller-Fleckenstein, B. Biesinger, A. Ensser, and H. Fickenscher.** 1993. Rhadinoviruses and t-cell transformation, p. S-23. Abstr. Int. Herpesvirus Meet.
35. **Gammelin, M., A. Altmuller, U. Reinhardt, J. Mandler, V. R. Harley, P. J. Hudson, W. M. Fitch, and C. Scholtissek.** 1990. Phylogenetic analysis of nucleoproteins suggest that human influenza A viruses emerged from a 19th-century avian ancestor. Mol. Biol. Evol. **7:**194–200.
36. **Geng, Y., B. Chandran, S. F. Josephs, and C. Wood.** 1992. Identification and characterization of a human herpesvirus 6 gene segment that *trans* activates the human immunodeficiency virus type 1 promoter. J. Virol. **66:**1564–1570.
37. **Gompels, U. A., M. A. Craxton, and R. W. Honess.** 1988. Conservation of gene organization in the lymphotropic herpesviruses, herpesvirus saimiri and Epstein-Barr virus. J. Virol. **62:**757–767.
38. **Gray, M. W.** 1992. The endosymbiont hypothesis revisited. Int. Rev. Cytol. **141:**233–373.
39. **Griffin, A. M.** 1991. The nucleotide sequence of the glycoprotein gB gene of infectious laryngotracheitis virus: analysis and evolutionary relationship to the homologous gene from other herpesviruses. J. Gen. Virol. **72:**393–398.
40. **Hayward, G. S., R. L. Jacob, S. C. Wadsworth, and B. Roizman.** 1975. Anatomy of herpes simplex virus DNA: evidence of four populations of molecules that differ in the relative orientations of their long and short components. Proc. Natl. Acad. Sci. USA **72:**4243–4247.
41. **Henikoff, S., and J. G. Henikoff.** 1992. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA **89:**10915–10919.
42. **Hollander, M., and D. A. Wolfe.** 1973. Nonparametric statistical methods. John Wiley & Sons, New York.
43. **Honess, R. W.** 1984. Herpes simplex and "the Herpes complex": diverse observations and a unifying hypothesis. J. Gen. Virol. **65:**2077–2107.
44. **Honess, R. W., W. Bodemer, K. R. Cameron, H.-H. Niller, and B. Fleckenstein.** 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.
45. **Honess, R. W., U. A. Gompels, B. G. Barrell, M. Craxton, K. R. Cameron, R. Staden, Y.-N. Chang, and G. S. Hayward.** 1989. Deviations from expected frequencies of CpG dinucleotides in herpesvirus DNAs may be diagnostic of differences in the states of their latent genomes. J. Gen. Virol. **70:**837–855.

45a. **Hsu, D. H., R. de Waal Malefyt, D. F. Fiorentino, M. N. Dang, P. Vieira, J. de Vries, H. Spits, T. R. Mosmann, and K. W. Moore.** 1990. Expression of interleukin-10 activity by Epstein-Barr virus protein BCRF1. Science **250:**830–832.

46. **Jansson, A., M. Masucci, and L. Rymo.** 1992. Methylation of discrete sites within the enhancer region regulates the activity of the Epstein-Barr virus *Bam*HI W promoter in Burkitt lymphoma lines. J. Virol. **66:**62–69.
47. **Jones, D., R. Isfort, R. Witter, R. Kost, and H.-J. Kung.** 1993. Retroviral insertions into a herpesvirus are clustered at the junctions of the short repeat and short unique sequences. Proc. Natl. Acad. Sci. USA **90:**3855–3859.
48. **Jones, D. T., W. R. Taylor, and J. M. Thomton.** 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. **8:**275–282.
49. **Jukes, T. H., and V. Bhushan.** 1986. Silent nucleotide substitutions and G+C content of some mitochondrial and bacterial genes. J. Mol. Evol. **24:**39–44.
50. **Karlin, S., B. E. Blaisdell, and G. A. Schachtel.** 1990. Contrasts in codon usage of latent versus productive genes of Epstein-Barr virus: data and hypotheses. J. Virol. **64:**4264–4273.
51. **Karlin, S., C. Burge, and A. M. Campbell.** 1992. Statistical analyses of counts and distributions of restriction sites in DNA sequences. Nucleic Acids Res. **6:**1363–1370.
52. **Karlin, S., and L. R. Cardon.** Computational DNA sequence analysis. Annu. Rev. Microbiol., in press.
53. **Karlin, S., R. Kennett, and B. Bonné-Tamir.** 1979. Analysis of biochemical genetic data on Jewish populations II. Am. J. Hum. Genet. **31:**341–365.
54. **Klupp, B. G., H. Kern, and T. C. Mettenleiter.** 1992. The virulence-determining genomic *Bam*H1 fragment 4 of pseudorabies virus contains genes corresponding to the U_L15 (partial), U_L18, U_L19, U_L20, and U_L21 genes of herpes simplex virus and a putative origin of replication. Virology **191:**900–908.

54a. **Kondo, K., T. Kondo, T. Okuno, M. Takahashi, and K. Yamanishi.** 1991. Latent human herpesvirus 6 infection of human monocytes/macrophages. J. Gen Virol. **72:**1401–1408.

54b. **Kondo, K., H. Nagafuji, A. Hata, C. Tomomori, and K. Yamanishi.** 1993. Association of human herpesvirus 6 infection of the central nervous system with recurrence of febrile convulsions. J. Infect. Dis. **167:**1197–1200.

55. **Ladunga, I., and S. Karlin.** Unpublished observations.
56. **Lake, J. A.** 1987. Determining evolutionary distances from highly diverged nucleic acid sequences: operator metrics. Mol. Biol. Evol. **26:**59–73.
57. **Lake, J. A.** Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. Proc. Natl. Acad. Sci. USA, in press.

57a. **Lawrence, G. L., M. Chee, M. A. Craxton, U. A. Gompels, R. W. Honess, and B. G. Barrell.** 1990. Human herpesvirus 6 is closely related to human cytomegalovirus. J. Virol. **64:**287–299.

58. **Li, W. H., and D. Graur.** 1991. Fundamentals of molecular evolution. Sinauer Press, Sunderland, Mass.
59. **Li-Fu, H., J. Minarovitz, C. Shi Long, B. Contreras-Salazar, L. Rymo, K. Falk, G. Klein, and I. Ernberg.** 1991. Variable expression of latent membrane protein in nasopharyngeal carcinoma can

be related to methylation status of the Epstein-Barr virus BNLF1 5′ flanking region. J. Virol. **65:**1558–1567.

60. **Lindquester, G. J., and P. E. Pellett.** 1991. Properties of the human herpesvirus 6 strain Z29 genome: G+C content, length, and presence of variable-length directly repeated terminal sequence elements. Virology **182:**102–110.
61. **Marchini, A., B. Tomkinson, J. I. Cohen, and E. Kieff.** 1991. BHRF1, the Epstein-Barr virus genes with homology to Bc12, is dispensable for B-lymphocyte transformation and virus replication. J. Virol. **65:**5991–6000.
62. **Martin, M. E. D., J. Nicholas, B. J. Thomson, C. Newman, and R. Honess.** 1991. Identification of a transactivating function mapping to the putative immediate-early locus of human herpesvirus 6. J. Virol. **65:**5381–5390.
63. **Minarovitz, J., S. Minarovitz-Kormuta, B. Ehlin-Henriksson, K. Falk, G. Klein, and I. Emberg.** 1991. Host cell phenotype dependent methylation pattern of Epstein-Barr virus DNA. J. Gen. Virol. **72:**1591–1599.
64. **McClelland, M., R. Jones, Y. Patel, and M. Nelson.** 1987. Restriction endonucleases for pulsed field mapping of bacterial genomes. Nucleic Acids Res. **15:**5985–6005.
65. **McGeoch, D. J.** 1989. The genomes of the human herpesviruses: contents, relationships, and evolution. Annu. Rev. Microbiol. **43:**235–265.
66. **McGeoch, D. J.** 1990. Evolutionary relationships of virion glycoprotein genes in the S regions of alphaherpesvirus genomes. J. Gen. Virol. **71:**2361–2367.
67. **McGeoch, D. J.** 1992. Molecular evolution of large DNA viruses of eukaryotes. Semin. Virol. **3:**399–409.

67a. **McGeoch, D. J.** 1993. Origins of the herpesviruses, p. S-26. Abstr. Int. Herpesvirus Meet.

68. **Nei, M.** 1987. Molecular evolutionary genetics. Columbia University Press, New York.

68a. **Neote, K., D. DiGregorio, J. Y. Mak, R. Horuk, and T. J. Schall.** 1993. Molecular cloning, functional expression, and signaling characteristics of a C-C chemokine receptor. Cell **72:**415–425.

68b. **Olsen, G. J., and C. R. Woese.** 1993. Ribosomal RNA: a key to phylogeny. FASEB J. **7:**113–123.

69. **Quinlan, M. P., and D. M. Knipe.** 1983. Nuclear localization of herpesvirus proteins: potential role for the cellular framework. Mol. Cell. Biol. **3:**315–324.
70. **Roizman, B.** 1979. The structure and isomerization of herpes simplex virus genomes. Cell **16:**487–494.
71. **Roizman, B.** 1979. The organization of herpes simplex virus genomes. Annu. Rev. Genet. **13:**25–57.
72. **Ross, L. J. N., and M. M. Binns.** 1991. Properties and evolutionary relationships of the Marek's disease virus homologues of protein kinase, glycoprotein D and glycoprotein I of herpes simplex virus. J. Gen. Virol. **72:**939–947.
73. **Roughgarden, J.** 1979. Theory of genetics and evolutionary ecology: an introduction. MacMillan, New York.
74. **Schachtel, G. A., P. Bucher, E. S. Mocarski, B. E. Blaisdell, and S. Karlin.** 1991. Evidence for selective evolution in codon usage in conserved amino acid segments of human alphaherpesvirus proteins. J. Mol. Evol. **33:**483–494.
75. **Selker, E. U.** 1990. DNA methylation and chromatin structure: a view from below. Trends Biochem. Sci. **15:**103–107.
76. **Stewart, C.-B.** 1993. The powers and pitfalls of parsimony. Nature (London) **361:**603–607.
77. **Swofford, D. L.** 1992. PAUP: phylogenetic analysis using parsimony, version 3.0. Center for Biochemistry, Illinois Natural History Survey, Champaign, Ill.
78. **Swofford, D. L., and G. J. Olsen.** 1990. Phylogeny reconstruction in molecular systematics, p. 411–501. *In* D. M. Hillis and C. Moritz (ed.), Molecular systematics. Sinauer Press, Sunderland, Mass.

78a. **Takahashi, K., S. Sonoda, K. Higashi, T. Kondo, H. Takahashi, M. Takahashi, and M. Yamanishi.** 1989. Predominant CD4 T-lymphocyte tropism of human herpesvirus 6-related virus. J. Virol. **63:**3161–3163.

79. **Tazi, J., and A. Bird.** 1990. Alternative chromatin structure at CpG islands. Cell **60:**909–920.
80. **Telford, E. A. R., M. J. Studdert, C. T. Agius, M. S. Watson, H. C. Aird, and A. J. Davison.** 1993. Equine herpesviruses 2 and 5 are γ-herpesviruses. Virology **195:**492–499.
81. **Telford, E. A. R., M. S. Watson, K. McBride, and A. J. Davison.** 1992. The DNA sequence of equine herpesvirus-1. Virology **189:**304–316.
82. **Teo, I. A., B. E. Griffin, and M. D. Jones.** 1991. Characterization of the DNA polymerase gene of human herpesvirus 6. J. Virol. **65:**4670–4680.
83. **Thiry, E., M. Bublot, J. Dubuisson, M.-F. Van Bressem, A.-S. Lequarre, P. Lomonte, A. Vanderplasschen, and P. P. Pastoret.** 1992. Molecular biology of bovine herpesvirus type 4. Vet. Microbiol. **33:**79–92.

83a. **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.

84. **Trifonov, E. N.** 1991. DNA in profile. Trends Biochem. Sci. **12:**467–470.
85. **Trimble, J. J., C. S. Murthy, A. Bakker, R. Grassmann, and R. C. Desroisiers.** 1988. A gene for dihydrofolate reductase in a herpesvirus. Science **239:**1145–1147.
86. **Wang, T. S., S. W. Wong, and D. Korn.** 1989. Human DNA polymerase alpha: predicted functional domains and relationships with viral DNA polymerases. FASEB J. **3:**14–21.
87. **Wolffe, A.** 1991. Chromatin structure and function. Academic Press, London.