VIROLOGY 244, 119–132 (1998)
ARTICLE NO. VY989105

The DNA Sequence of the RK Strain of Human Herpesvirus 7

A. George Megaw,* Debora Rapaport,† Boaz Avidor,‡ Niza Frenkel,† and Andrew J. Davison*[1]

*MRC Virology Unit, Institute of Virology, Church Street, Glasgow G11 5JR, United Kingdom; †Department of Cell Research and Immunology, Tel-Aviv University, Tel-Aviv 69978, Israel; and ‡The Bernard Pridan Laboratory for Molecular Biology of Infectious Diseases, Ichilov Hospital, Sourasky Medical Center, Tel-Aviv 64239, Israel

Received December 19, 1997; returned to author for revision February 6, 1998; accepted February 20, 1998

The complete DNA sequence of human herpesvirus-7 (HHV-7) strain RK was determined following direct cloning of virion DNA fragments into a sequencing vector. The sequence was compared with the previously published complete sequences of HHV-7 strain JI and human herpesvirus-6 (HHV-6) strain U1102. Despite a very close relationship between the two HHV-7 strains, differences are apparent in regions containing tandem reiterations, particularly in the "telomeric" reiterations located near the termini of the large direct repeat at the genome ends, and in a total of 179 additional positions distributed throughout the genome (i.e., about one nucleotide difference per kbp). This extent of divergence implies that the two strains arose from an ancestral virus several thousands of years ago. Differences that affect coding potential do not cluster in particular protein-coding regions, indicating that specific HHV-7 genes have not been measurably subject to unusual evolutionary pressures since divergence. Reassessments of genetic content indicated that the HHV-7 genome contains 84 different genes, whereas the HHV-6 genome contains 85. All HHV-7 genes but 1 have direct HHV-6 counterparts, and all but 2 HHV-6 genes have HHV-7 homologues. Sequence comparisons between HHV-7 and HHV-6 provided evidence that the protein-coding regions of 11 genes are expressed by splicing.

INTRODUCTION

Human herpesvirus-7 (HHV-7) was first isolated in 1989 at the National Institutes of Health (Bethesda, MD) by Frenkel *et al.* (1990) after activation of $CD4^+$ T cells purified from the peripheral blood of a healthy individual (initials RK). This isolate was designated HHV-7(RK), following the rules of the Herpesvirus Study Group of the ICTV (Roizman *et al.,* 1981). The virus was shown to be related to the two variants of human herpesvirus-6 (HHV-6A and -6B) by DNA hybridization, and yet exhibited restriction endonuclease patterns that are distinct from those of both variants. Subsequently, other strains were isolated, including HHV-7(JI), which was obtained by Berneman *et al.* (1992b) at the National Cancer Institute (Bethesda, MD) from the peripheral blood mononuclear cells of a patient suffering from chronic fatigue syndrome. Initial DNA sequence data from HHV-7(JI) reinforced the view that HHV-7 and HHV-6 are closely related but distinct members of the Betaherpesvirinae (Berneman *et al.,* 1992a). HHV-6 and HHV-7 are ubiquitous in the human population (Okuno *et al.,* 1989; Wyatt *et al.,* 1991), and one of the HHV-6 variants (HHV-6B) has been proposed as the causative agent of a childhood disease, exanthem subitum (Yamanishi *et al.,* 1988). HHV-7 has no proven involvement in any disease. A potential role in a proportion of exanthem subitum cases has been suggested (Tanaka *et al.,* 1994).

The genome sequence of HHV-7(JI) was derived by Nicholas (1996) from plasmid and bacteriophage λ clones. It yielded a detailed picture of the organization of HHV-7 genes and their relationships to genes of other sequenced herpesviruses, particularly HHV-6A strain U1102 (Gompels *et al.,* 1995) and human cytomegalovirus (HCMV) strain AD169 (Chee *et al.,* 1990). The HHV-7(JI) genome is 144,861 bp in size, and consists of a unique region (U) of 133,233 bp flanked by a large direct repeat (DR) of 5814 bp. A total of 101 open reading frames (ORFs) potentially encoding proteins were identified, 8 of them present in both copies of DR. Of these, 11 lack counterparts in HHV-6.

The HHV-6A(U1102) genome is 159,321 bp in size and has a similar structure to that of HHV-7, U and DR being 143,147 and 8087 bp in size, respectively. Of the 113 ORFs (8 present in both copies of DR) listed by Gompels *et al.* (1995), 22 lack counterparts in HHV-7(JI). Taking into account possible splicing and excluding some ORFs thought unlikely to encode proteins, the HHV-6 genome was considered to contain 102 separate genes. The majority of HHV-6 (and HHV-7) genes have counterparts in HCMV, and about 40 have counterparts in all sequenced mammalian herpesviruses.

In this paper, we report the complete DNA sequence of HHV-7(RK). We present reassessments of the genetic contents of HHV-7 and HHV-6 derived by comparing the HHV-7(RK), HHV-7(JI), and HHV-6A(U1102) sequences.

[1] To whom reprint requests should be addressed. Fax: +44 141 337 2236. E-mail: a.davison@vir.gla.ac.uk.

0042-6822/98 $25.00

Evidence for splicing in several genes is of special interest. Readers should note that these studies build on the extensive sequence interpretations of Gompels *et al.* (1995) and Nicholas (1996), and reference to these analyses is implicit throughout this paper. Also, we have used citations sparingly, since very extensive bibliographies were provided in these earlier publications.

RESULTS AND DISCUSSION

Viral origin of the sequenced DNA

HHV-7(RK) genomic DNA was isolated from capsids on the principle that encapsidated viral DNA, unlike cellular or nonencapsidated viral DNA, is protected from digestion by DNase I. As the sequencing protocol was based on alignment of the sequences of a large number of random fragments generated by sonication of this DNA, it was important to verify the purity of the preparations (data not shown). Each of the four preparations analysed was found to contain a few micrograms of genome-length DNA plus a much larger amount of fragments smaller than genome size (up to several kbp) or 100 bp, or both. The provenance of these fragments was not determined, and they may represent residual viral or cellular DNA that was degraded during DNase digestion but remained associated with capsids.

Genome-length DNA was selectively precipitated using polyethylene glycol from a DNA preparation that was heavily contaminated with fragments smaller than 100 bp. The proportion of random sequences obtained using this material that contributed to the final genomic sequence indicated that the 97% originated from the RK genome.

Genome features

The final database obtained by entry of all random sequences was edited by careful reference to autoradiographs. The sequence was determined an average of 8.2 times per nucleotide, and 97% was determined on both strands. Regions in which data were obtained for only one strand were examined particularly closely during editing. Ambiguities throughout the sequence were then resolved by analyses of specific templates from both strands using the universal or custom primers. At this stage, the RK sequence was compared with the published sequence of HHV-7(JI). Positions at which the RK sequence differed from that of JI were checked by reference to autoradiographs, and confirmatory data were obtained when necessary.

Sequences containing the genome termini were identified in the database inasmuch as they have one unique end, unlike other fragments generated randomly from the genome. Most clones from the left terminus ended in six C residues, but the possibility of heterogeneity was suggested by the presence of fewer C residues in some clones. As fragments were end-repaired prior to ligation, no information was obtained on unpaired nucleotides at the termini. The genome termini of another HHV-7 strain (R-2), also sequenced from end-repaired DNA, are identical (Secchiero *et al.,* 1995). Sequences at the termini of JI have not been determined directly. However, concatemeric DNA from HHV-7-infected cells contains head-to-tail copies of the genome, and the DR–DR junction thus corresponds to a fused version of the termini (including any unpaired residues normally present at the termini). The sequence of this region in JI concatemers is variable, containing four to six C residues in the region corresponding to the left terminus plus, in some instances, one or two additional nucleotides at the junction itself (Secchiero *et al.,* 1995). Again, this suggests that the tract of C residues at the left terminus is variable in length. The right terminus of the published JI genome sequence is identical to that of RK, but the left terminus is one base pair longer, containing seven C residues. The origin of this additional residue is unclear.

No evidence was obtained from the random sequence data to indicate that there are any differences between the two copies of DR (also termed TR by Frenkel and Roffman (1996)). As a consequence, the final database contained only one copy. Sequences at the genome termini were used to identify the DR–U and U–DR junctions and thus to define the DR sequence. The genome sequence was reconstructed to represent U (133,012 bp) flanked on each side by a copy of DR (10,034 bp). The total size of the sequence is, therefore, 153,080 bp.

Strain variation

Four regions in the RK genome contain head-to-tail reiterations of short sequence elements. Their locations are shown in Fig. 1. Two (R1 and R2) are located in U, and two (telomeric reiterations T1 and T2) are located near the ends of DR. Reiterations are located at similar positions in JI, but each is different in size and sequence to its RK counterpart. In RK, R1 comprises three copies of an 84-bp element and a partial copy of 68 bp, with two mismatched nucleotides. In JI, the third copy of the element is only 81 bp long, differing from the other two elements in several positions, and the partial copy is shorter (66 bp). R2 consists of an array of related 105-bp elements. RK has 15 copies plus a partial copy of 26 bp, and JI has 17 copies plus the partial copy. The RK and JI sequences are similar towards the ends of the array, but differ in the central region.

The basic element of T1 and T2 (TAACCC) is related to human telomeric sequences. These reiterations consist of arrays of this element interspersed with 34 types of related elements, most also hexanucleotides but some larger. Sequences for these regions were derived from the random sequence data with difficulty. The range of different element types aided the analysis, but their pres-

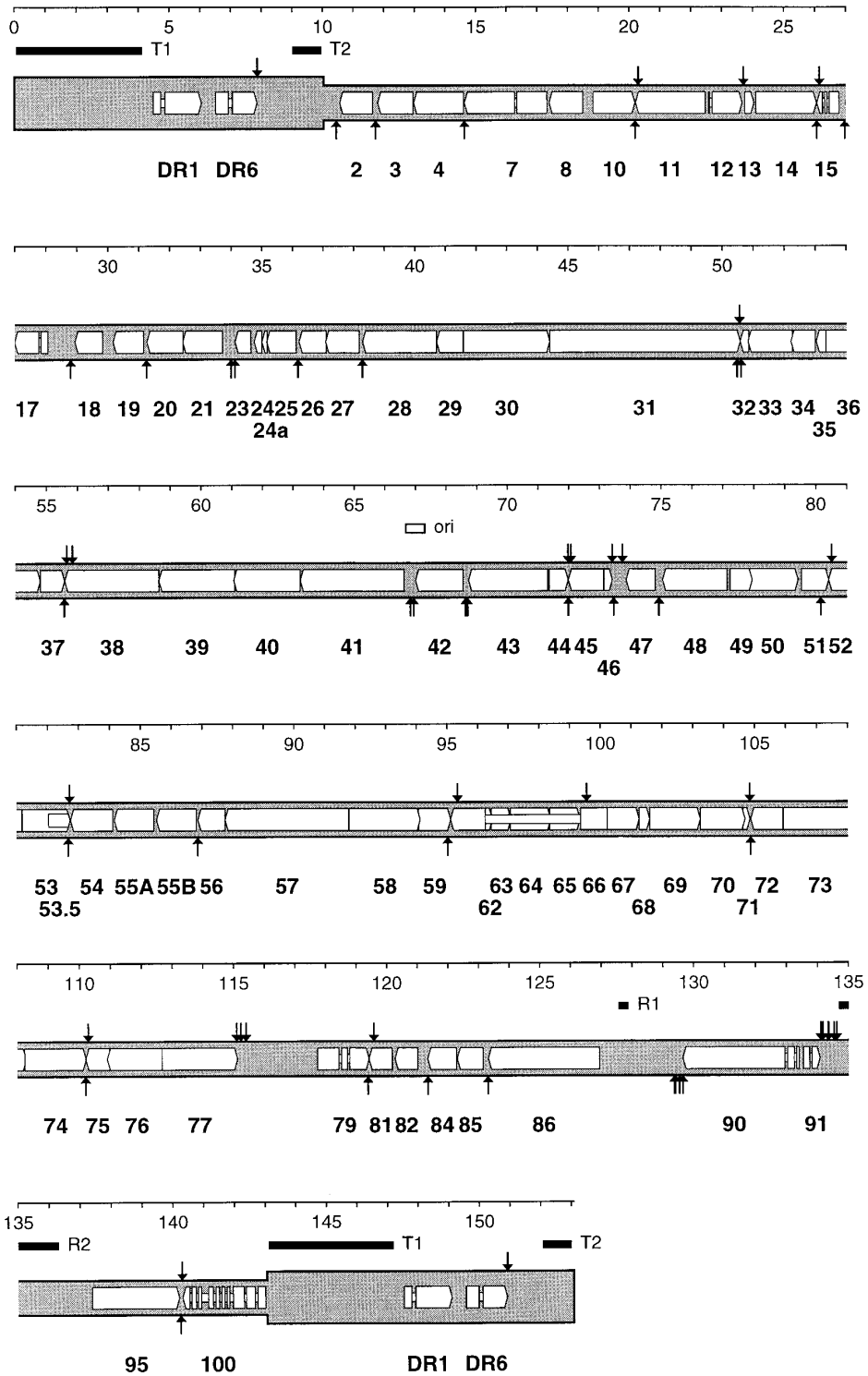

FIG. 1 Predicted HHV-7 gene arrangement. The genome is shaded, the thinner and thicker portions denoting the unique region and terminal direct repeats, respectively. The scale is in kbp. Protein-coding regions (DR1, DR6, U2-U100) are shown as open arrows above the gene nomenclature; the U prefix has been omitted. ORFs predicted to be expressed as spliced mRNAs are connected by open horizontal bars. Candidate polyadenylation signals (AATAAA, ATTAAA, or AGTAAA) are indicated by vertical arrows in the appropriate strand. Reiterated sequences (R1, R2, T1, and T2) are shown as filled rectangles above the genome. The minimal fragment shown to act as an origin of DNA replication (ori) is indicated (van Loon *et al.,* 1997).

ence at two separate loci was a complicating factor. An acceptable solution for T2 was obtained, but that for T1 was ultimately unsatisfactory, as described below. The large size and reiterated structure of T1 preempted sequencing it from single templates using the universal or custom primers.

T2 is the shorter reiteration, consisting of 148 copies of the telomeric element. The *Bam*HI fragment at the right genome terminus contains T2, and a close correspondence in size is evident between the RK fragment predicted from the sequence (3811 bp) and that detected in genomic DNA by DNA hybridization (3.8 kbp; Singer and Frenkel, 1997).

The solution derived for T1 contains 663 copies of the telomeric element. It is probably incorrect, however, since the size of the predominant *Bam*HI fragment identified by DNA hybridization at the left terminus of genomic DNA is 1.7 kbp smaller than that predicted from the sequence (Singer and Frenkel, 1997). Moreover, a few random fragments could not be accommodated in the deduced sequence. This may reflect either the intractability of this region or sequence heterogeneity occurring in the genome or in M13 recombinants. Variability in the genome itself is supported by the observation that the terminal *Bam*HI fragment identified by Singer and Frenkel (1997) is heterogeneous in size. Therefore, it must be emphasized that although the size of the RK genome as derived from the sequence is 153,080 bp, the actual size of the major species is likely to be a little less than 150 kbp.

Bacterial clones containing T1 and T2 from JI contain 20 and 106 copies of the telomeric element, respectively (Secchiero *et al.,* 1995). Both are thus smaller than in RK, the former markedly so. The corresponding sequences in JI genomic DNA are likely to be different, however, since they vary in size in different plasmid clones (Ruvulo *et al.,* 1996). In support of the view that HHV-7 does not have a unique genome size, we note that Frenkel and Roffman (1996) reported that the genomes of different HHV-7 strains range in size from 140 to 150 kbp, and suggested that the main cause of variation involves the telomeric reiterations.

The presence of telomeric reiterations in the HHV-7 and HHV-6 genomes is of unknown significance. Published speculations include roles in genome cleavage, regulation of gene expression, segregation during cell division, latent-phase viral DNA replication, and site-specific recombination into the host genome. Whatever their function, if any, it appears that the telomeric reiterations are the most disparate regions in the RK and JI genomes, and are responsible for almost all of the size difference between the two sequences.

Many positions outside the reiterations differ in the two HHV-7 sequences, most as nucleotide substitutions but some as small insertions or deletions. In determining the origins of these differences, the RK sequence was examined carefully in these regions, obtaining new data when necessary in order to exclude errors. Next, inferred errors in the JI sequence were identified as insertions or deletions in coding regions that result in frameshifts causing loss of amino acid sequence similarity to HHV-6. Only four errors were noted, two in ORF U38 (2-bp deletion at 57,718 and a 1-bp deletion at 57,733) and two in ORF U53 (1-bp deletion at 81,361 and a 2-bp deletion plus a 2-bp substitution at 81,380). Each causes a local shift from and then back into the correct reading frame. In principle, some of the remaining differences may also result from errors in the JI sequence. There is, however, no evidence to support this, and all other disparities were classified as reflecting genuine strain variation. The types and distribution of differences are listed in Table 1. As anticipated, the majority of differences do not affect coding capacity. Moreover, differences that do affect coding potential are scattered throughout the genome, and do not appear to be clustered in particular genes. This indicates that no small subset of genes has been demonstrably subject to unusual evolutionary pressures since divergence of the two strains.

TABLE 1

Differences between the Genome Sequences of HHV-7(RK) and (JI)

Type of difference	Genome region		
	DR	U	Total[a]
Synonymous substitution	10	84	104
Nonsynonymous substitution	2	46	50
Multiple substitution[b]	1	2	4
Insertion/deletion[c]	5	11	21
Total	18	143	179

[a] U plus two copies of DR.
[b] Two or more consecutive nucleotide substitutions.
[c] One or a small number of nucleotides.

McGeoch and Cook (1994) demonstrated that the phylogeny of most members of the Alphaherpesvirinae is similar to that of their hosts. They concluded that these viruses have cospeciated with their hosts over periods up to 6×10^7 years, derived a timescale for evolution of the subfamily, and estimated an overall rate of sequence change of 1×10^{-7} synonymous substitutions per synonymous site per year in the gB gene. This value was claimed as an order of magnitude estimate only. It was considered compatible with estimates derived by analyses of herpes simplex virus type 1 isolates from human populations established over a much shorter period (10^5 years). On this basis, the presence of 84 synonymous substitutions in HHV-7 U (which, unlike DR, is not potentially subject to intragenomic homologous recombination) indicates that the lineages leading to RK and JI diverged from a common ancestor of the order of 10^4 years ago. In contrast, HHV-7 and HHV-6 are estimated to have diverged 5-6 $\times 10^7$ years ago (D.J. McGeoch, personal communication).

Genetic content

The criteria used to identify genes in a nucleotide sequence are to some extent arbitrary and inevitably

lead to false positives and false negatives. This applies even to criteria that imply numerical exactitude, such as minimum ORF length. Thus, a prediction of genetic content actually takes the form of a spectrum of probabilities, ranging from solid candidates (usually substantial ORFs with counterparts in other genomes) to marginal possibilities (usually smaller ORFs lacking counterparts). The effects of different sets of criteria are usually apparent in the lower reaches of the spectrum. Also, as data become available, specific factors occasionally have a dominant effect on individual candidates regardless of other considerations (experimental information on gene expression, for example).

In evaluating the genetic content of HHV-7 and HHV-6, we started with the previous predictions of coding capacity. Primary recognition was given to potential genes identified as ORFs that encode proteins with sequence counterparts in the other virus. Since the great majority of herpesvirus genes have individual polyadenylation signals close downstream from the ORF or are expressed as part of a 3′ coterminal family of ORFs with a polyadenylation signal close downstream from the most 3′ member, appropriate location of such signals was also an important factor. ORFs unique to each genome but substantial in size and possessing polyadenylation signals were then considered. Similarities to proteins encoded by other herpesviruses or other organisms, features diagnostic of membrane proteins, and potential splicing and transcript mapping data were also taken into account. Some previously identified ORFs suffer from combinations of features indicating that they are less likely to encode proteins: small size, lack of suitable initiation codons, significant overlap with other genes, presence of extensive reiterated sequences, absence of suitable polyadenylation signals, and lack of sequence similarity to other proteins. We consider that the status of these ORFs is sufficiently uncertain to exclude them as candidate genes at present. Codon usage or base preference analyses were not of strong predictive value in evaluating genetic content; in any case, they would have been of limited use for the small ORFs which make up most of the marginal category.

This rather stringent approach effectively resulted in a spectrum of probabilities for each virus that is truncated at its lower margin in comparison with those derived previously. We are aware that it is biased towards excluding genes that are present in only one of the two viruses (false negatives) rather than to inclusion of ORFs that do not encode proteins (false positives). Nevertheless, each candidate ORF was considered carefully in developing the schemes described below. Thus, we view the results as closer representations of the genetic contents of HHV-7 and HHV-6 than were previously available, both in terms of exclusion of marginal ORFs and in identification of potentially spliced genes (described below). One consequence is that the genetic complements of HHV-7 and HHV-6 appear more similar than previously proposed.

The deduced layout of HHV-7 genes is shown in Fig. 1, and details are listed in Table 2. The nomenclature developed for HHV-6 and extended to JI was employed for RK, except that several ORFs potentially expressed by splicing are identified only by the 5′-proximal ORF. Of the ORFs identified in the JI sequence that are absent from Fig. 1, parts of DR2, DR7, U5, U16, U17, U17ex, U60, U80, U89, H6, H8, U98, and U99 now form exons in spliced genes; U17ex spliced to U16 is termed U17. The remaining JI ORFs lacking counterparts in HHV-6 (H1, H2, H3, H4, U17a, H5, and H7) are considered unlikely to encode proteins for reasons listed above. Moreover, H3 and H4 are disrupted by insertions or deletions in RK.

As a result of this analysis, we consider that HHV-7 has 84 different genes, 2 (DR1 and DR6) present twice, giving a total of 86 genes in the genome. The most marginal candidate, U24a, is small and poorly conserved, but encodes a protein containing a highly hydrophobic domain in both HHV-7 and HHV-6. All proposed genes but one have direct HHV-6 counterparts. The exception, U55B, is a sizeable member of a 3′-coterminal family, and was shown previously to be related to an adjacent member, U55A. Both genes are thus related to HHV-6 U55. All proposed HHV-7 genes have candidate polyadenylation signals near the 3′ end of the ORF or 3′-coterminal family, as shown in Fig. 1.

Similar criteria applied to the HHV-6 sequence indicate that the genome has 85 different genes (87 total), all with appropriate polyadenylation signals except U2, which instead has a CATAAA element. Of the 22 HHV-6 ORFs identified by Gompels *et al.* (1995) that lack HHV-7 counterparts, only two are strong candidates for encoding complete proteins: U22, a potential membrane glycoprotein, and U94, a homologue of the adeno-associated virus type 2 *rep* gene. HHV-6 U87 now forms part of U86, owing to a sequencing error by Nicholas (1994) that was corrected by Nicholas (1996). Most ORFs lacking counterparts in HHV-7 (LT1, DR3, DR4, DR5, DR8, LJ1, U1, U6, U9, U61, U78, U83, U88, U92, U93, and RJ1) are considered unlikely to encode proteins. U12ex, U96, and U97 are predicted to form exons in spliced genes.

Where possible, amino acid sequence similarity between HHV-7 and HHV-6 was also used as a guide to locating initiation codons and possible splice sites. This resulted in modifications to the locations of some HHV-7 genes from those reported for JI, and details of these are incorporated into Table 2. Suggested modifications to the genes listed by Gompels *et al.* (1995) are given in Table 3; a few were included in publications that preceded completion of the HHV-6 sequence or were implied by Nicholas (1996).

About half of the genes in HHV-7 and HHV-6 have counterparts in all sequenced mammalian herpesviruses; these are termed core genes and are highlighted

TABLE 2

Features of HHV-7(RK) Genes

Gene[a]		Strand	Start[b]	Stop[c]	Codons	ID[d]	Properties
DR1	ex1	+	4485	4728	475	33	HCMV US22 gene family; exon 1 starts at 4466
	ex2		4863	6046			
DR6	ex1	+	6509	6926	406	59	HCMV US22 gene family
	ex2		7055	7857			
U2		-	11637	10558	359	48	HCMV US22 gene family
U3		-	12953	11799	384	49	HCMV US22 gene family
U4		-	14603	12975	542	59	Related to U7 exon 2
U7	ex1	-	17324	16348	>871	54	HCMV US22 gene family (exon 1); related to U4 (exon 2);
	ex2		16266	14628			exon 1 starts at 17326
U8		-	18483	17395	362	51	HCMV US22 gene family
U10		+	18829	20184	451	50	
U11		-	22470	20203	755	31	Structural phosphoprotein
U12	ex1	+	22566	22598	333	47	G protein-coupled receptor
	ex2		22689	23657			
U13		+	23742	24038	98	33	
U14		+	24106	26052	648	50	HCMV UL25/UL35 gene family
U15	ex1	-	26785	26469	191	68	
	ex2		26393	26321			
	ex3		26249	26064			
U17	ex1	-	28057	27841	330	53	HCMV US22 gene family; IE-B[f] transactivator
	ex2		27768	26993			
U18		-	29821	28934	295	44	IE-B membrane glycoprotein
U19		-	31142	30165	325	34	IE-B protein
U20		-	32431	31256	391	22	Probable membrane glycoprotein
U21		-	33714	32422	430	31	Probable membrane glycoprotein
U23		-	34638	34123	171	-	Probable membrane glycoprotein
U24		-	34992	34744	82	28	Contains a hydrophobic domain
U24A		-	35166	34996	56	25	Contains a hydrophobic domain
U25		-	36118	35156	320	47	HCMV US22 gene family
U26		-	37090	36209	293	30	
U27		-	38172	37078	364	68	Processivity subunit of replicative DNA polymerase; [*UL42*]
U28		-	40705	38285	806	47	Ribonucleotide reductase large subunit [UL39]
U29		-	41568	40708	286	53	Capsid protein; component of intercapsomeric triplex [UL38]
U30		+	41583	44399	938	46	Tegument protein [*UL37*]
U31		+	44400	50579	2059	46	Very large tegument protein [UL36]
U32		-	50848	50576	90	66	Capsid protein; located on tips of hexons [*UL35*]
U33		-	52262	50829	477	59	Virion protein
U34		-	52989	52213	258	56	Membrane-associated phosphoprotein [*UL34*]
U35		-	53340	53026	104	58	Role in DNA packaging [UL33]
U36		+	53339	54796	485	58	Role in DNA packaging [UL32]
U37		+	54798	55577	259	62	[UL31]
U38		-	58625	55584	1013	67	Catalytic subunit of replicative DNA polymerase [UL30]
U39		-	61093	58625	822	56	Envelope glycoprotein gB [UL27]
U40		-	63221	61056	721	56	Role in DNA packaging [UL28]
U41		-	66619	63224	1131	68	Single-stranded DNA-binding protein [UL29]
U42		-	68546	66996	516	56	Post-translational regulator of gene expression [UL54]
U43		-	71310	68725	861	61	Component of DNA helicase-primase complex; primase [UL52]
U44		+	71367	71978	203	58	[*UL51*]
U45		-	73122	71983	379	50	[UL50][g]
U46		+	73154	73414	86	52	Membrane protein [UL49A]
U47		-	74803	73862	313	23	
U48		-	77133	75041	690	39	Envelope glycoprotein gH; complexes with gL [UL22]
U49		+	77226	77945	239	52	[UL24]
U50		+	77761	79425	554	55	Role in DNA packaging [UL25]
U51		+	79527	80411	294	35	G protein-coupled receptor
U52		-	81172	80408	254	56	
U53		+	81180	82721	513	52	N-terminal protease domain acts in capsid maturation and is a capsid protein; C-terminal domain is the minor capsid scaffold protein [UL26]
U53.5		+	82029	82721	230	51	Major capsid scaffold protein [UL26.5]
U54		-	84100	82736	454	42	Virion transactivator
U55A		-	85431	84148	427	33	Related to U55B
U55B		-	86807	85515	430	21	Related to U55A
U56		-	87741	86860	293	65	Capsid protein; component of intercapsomeric triplex [UL18]
U57		-	91781	87744	1345	68	Major capsid protein; forms hexons and pentons [UL19]
U58		+	91793	94120	775	61	
U59		+	94068	95111	347	38	
U62		+	96247	96474	75	45	
U63		+	96446	97081	211	68	
U64		+	97059	98378	439	41	Role in DNA packaging; tegument protein [*UL17*]
U65		+	98341	99333	330	59	Tegument protein [UL16]

TABLE 2—*Continued*

Gene[a]		Strand	Start[b]	Stop[c]	Codons	ID[d]	Properties
U66	ex1	–	100215	99352	663	72	Role in DNA packaging; putative terminase [UL15]
	ex2		96235	95108			
U67		+	100214	101254	346	52	[*UL14*]
U68		+	101254	101598	114	48	
U69		+	101601	103241	546	53	Serine-threonine protein kinase; tegument protein [UL13]
U70		+	103243	104685	480	52	Deoxyribonuclease; role in maturation/packaging of DNA [UL12]
U71		+	104622	104843	73	53	Myristylated tegument protein [UL11]
U72		–	105906	104866	346	59	Envelope glycoprotein gM; role in virion envelopment [UL10]
U73		+	105923	108286	787	58	Origin-binding protein; helicase [UL9]
U74		+	108237	110216	659	41	Component of DNA helicase-primase complex [*UL8*]
U75		–	110973	110203	256	45	[UL7]
U76		–	112819	110897	640	59	Minor capsid protein; role in DNA packaging [UL6]
U77		+	112665	115127	820	75	Component of DNA helicase-primase complex; helicase [UL5]
U79	ex1	+	117733	118429	506	42	Probable role in DNA replication
	ex2		118522	118712			
	ex3		118788	119420			
U81		–	120179	119415	254	58	Uracil-DNA glycosylase [UL2]
U82		–	121009	120269	246	38	Envelope glycoprotein gL; complexes with gH [UL1]
U84		–	122271	121339	310	42	
U85		–	123141	122299	280	37	Probable membrane glycoprotein; related to OX-2
U86		–	126934	123317	1205	29	IE-A[h] protein
U90	ex1	–	133408	133323	1199	28	IE-A transactivator; exon 1 starts at 133434
	ex2		133244	133033			
	ex3		132949	129648			
U91	ex1	+	133539	133740	153	27	Probable membrane glycoprotein
	ex2		133817	134076			
U95		+	137378	140201	940	25	HCMV US22 gene family
U100	ex1	–	142997	142753	603	28	Envelope glycoprotein gp105; exon 1 starts at 143000
	ex2		142658	142377			
	ex3		142303	141947			
	ex4		141826	141749			
	ex5		141672	141565			
	ex6		141494	141381			
	ex7		141298	141144			
	ex8		140914	140815			
	ex9		140742	140599			
	ex10		140532	140304			
DR1	ex1	+	147531	147774	475	33	HCMV US22 gene family; exon 1 starts at 147512
	ex2		147909	149092			
DR6	ex1	+	149555	149972	406	59	HCMV US22 gene family
	ex2		150101	150903			

[a] Protein-coding exons (ex) are listed. Genes with counterparts in all mammalian herpesviruses are shaded.

[b] First exons, from first nucleotide of first complete codon (U7) or initiation codon (other ORFs). Subsequent exons, from first nucleotide.

[c] To last nucleotide of stop codon or exon.

[d] Percentage identical amino acid residues to the HHV-6 counterpart as determined by Gap at default values; U23 proteins did not align at these settings.

[e] Properties derived from the current analysis and summaries in Gompels *et al.* (1995), Nicholas (1996), and numerous other herpesvirus genome sequence papers. For genes with counterparts in all mammalian herpesviruses, the HSV-1 nomenclature is given in square parentheses; where identification is based on positional data, the HSV-1 gene is shown in italics.

[f] Immediate early B locus.

[g] Related to HSV-1 UL50 encoding deoxyuridine triphosphatase, but probably lacking enzymatic function.

[h] Immediate early A locus.

in Table 2. The previous studies pointed out that most noncore genes are located towards the genome termini, and many have counterparts in HCMV (Chee *et al.,* 1990) or murine cytomegalovirus (MCMV; Rawlinson *et al.,* 1996). Several of these genes belong to the US22 family, some members of which appear to be involved in gene regulation.

Splicing

Drawing largely on data from other herpesviruses (especially HCMV) and in some cases on transcript mapping data, previous work indicated that splicing may occur in some HHV-6 genes. However, the detailed situation was unclear. Splice sites were predicted for U66 (Lawrence *et al.,* 1990) and U17, U18, and U20 (Nicholas and Martin, 1994), and splicing patterns were deduced from transcript mapping data for U90, U91, and U100 (Schiewe *et al.,* 1994; Pfeiffer *et al.,* 1995). Subsequently, Gompels *et al.* (1995) listed sites only for HHV-6 U12 and U17, with three of the four sites apparently incorrect, as described below. Nicholas (1996) indicated that splicing may occur in certain HHV-7 genes, but predicted splice

TABLE 3

Proposed Modifications to HHV-6A(U1102) Genes

Gene[a]		Strand	Start[b]	Stop[c]	Codons	Reason for change
DR1	ex1	+	501	759	689	Putative splicing; exon 1 starts at 500
	ex2		843	2653		
DR6	ex1	+	4725	5028	395	Putative splicing
	ex2		5837	6720		
U7	ex1	−	15921	14948	>872	Putative splicing; exon 1 starts at 15923
	ex2		14858	13214		
U12	ex1	+	21680	21712	351	Correction of splice sites
	ex2		21790	22812		
U14		+	23331	25145	604	5′ truncation[d]
U15	ex1	−	25992	25676	191	Putative splicing
	ex2		25602	25530		
	ex3		25364	25179		
U17	ex1	−	27349	27121	334	Definition of splice sites
	ex2		27034	26259		
U24A		−	35847	35674	57	Not previously identified
U27		−	38903	37797	368	5′ truncation
U30		+	42325	45132	935	5′ truncation
U44		+	73470	74087	205	5′ truncation
U47		−	77768	75912	618	5′ truncation
U53.5		+	85133	85867	244	Definition of coding region
U66	ex1	−	102486	101614	666	Definition of splice sites
	ex2		98415	97288		
U67		+	102485	103519	344	5′ truncation
U76		−	115257	113317	646	5′ truncation
U79	ex1	+	120164	120794	474	Definition of splice sites
	ex2		120891	121087		
	ex3		121170	121766		
U86		−	130044	125989	1351	Error correction
U90	ex1	−	136112	136054	941	Definition of splice sites; exon 1 starts at 136150
	ex2		135965	135772		
	ex3		135664	133092		
U91	ex1	+	136267	136477	153	Definition of splice sites
	ex2		136580	136830		
U100	ex1	−	150282	149873	656	Definition of splice sites; exon 1 starts at 150295
	ex2		149771	149490		
	ex3		149081	148746		
	ex4		148628	148551		
	ex5		148454	148347		
	ex6		148255	148142		
	ex7		148055	147895		
	ex8		147383	147374		
	ex9		147223	147095		
	ex10		146984	146642		
DR1	ex1	+	151735	151993	689	Putative splicing; exon 1 starts at 151734
	ex2		152077	153887		
DR6	ex1	+	155959	156262	395	Putative splicing
	ex2		157071	157954		

[a] Protein-coding exons (ex) are listed. Coordinates include an extra G residue at 128132 (in U86) as indicated by Nicholas (1996).
[b] First exons, from first nucleotide of first complete codon (U7) or initiation codon (other ORFs). Subsequent exons, from first nucleotide.
[c] To last nucleotide of stop codon or exon.
[d] As determined by aligning N-terminal regions of homologous proteins.

sites only for U66 and U17. In reassessing this aspect of HHV-7 gene organization, we found substantial support from sequence comparisons for splicing in nine genes in addition to U66 and U17 (U7, U15, DR1, DR6, U12, U79, U90, U91, and U100), and were able to predict splicing patterns.

Appropriate parts of the RK and HHV-6 DNA sequences were aligned with conceptual translation products in all six reading frames, and splicing patterns were identified by visual comparison of the amino acid sequences at corresponding locations in the two genomes. In instances where similarity between potential protein-

TABLE 4

Predicted Splice Sites in the HHV-7 and HHV-6 Genomes

RK ORF		JI ORF	HHV-7 acceptor	HHV-7 donor	HHV-6 acceptor	HHV-6 donor
consensus			YYYYYYYYYYYYNYAGG	MAGGTRAGT	YYYYYYYYYYYYNYAGG	MAGGTRAGT
DR1	exon 1	DR1	TCTCTTCTATCACAGA	CATGTAAGC	CTATTCTTACTCTAGG	TATGTGAGT
	exon 2	DR2	TTTGCTCTATCGCAGG	None	TTTGGTTTCCCTCAGG	None
DR6	exon 1	DR6	None	GCGGTGAGT	None	GCGGTGAGT
	exon 2	DR7	TCTACATCCCGGCAGC	None	CGCGTCCCATCACAGC	None
U7	exon 1	U7	AAGTTAATTTTGCAGA	TAGGTATGT	TTATGTTTCAAACAGA	CAGGTGGGT
	exon 2	U5	GATGTTCTTTTTCAGA	None	TTTTTCTTACAACAGG	None
U12	exon 1	—	None	CTGGTATGA	None	CTGGTAAGT
	exon 2	U12	AACTTTTTTTCACAGC	None	CAATATCTTTAATAGC	None
U15	exon 1	U15	None	ACGGTGAGT	None	AAGGTGGGT
	exon 2	—	CTCTTTTATTTTCAGG	GCGGTAAGA	GTTTTTCTTTTACAGG	GCGGTGAGT
	exon 3	—	TTTTTTTTTCTTTAGT	None	TGTTTTTTTTTTAGT	None
U17	exon 1	U17	None	TATGTAAGT	None	TGTGTAAGT
	exon 2	U16	TTGTTGTTTTCATAGG	None	TTCCTTTTAAACAAGG	None
U66	exon 1	U66	None	CACGTAAGT	None	CACGTAAGT
	exon 2	U60	TCATTTTCTTCTCAGA	None	TCATTCCCCTCTCAGA	None
U79	exon 1	U79	None	AAGGTTAGT	None	ATGGTAAAT
	exon 2	H6	AACATGTTTTCTTAGA	CAGGTGGGT	TTGTTGCAATTTCAGA	CAGGTGGGT
	exon 3	U80	GTTTCTTTCTTTTAGG	None	ATCTTTTTATTTTAGG	None
U90	exon 1	—	GGTTTGTTATTGTAGG	TGAGTAGGT	TTCATTGGCTATCAGC	AGAGTAAGT
	exon 2	U90	TAAATTTTATTACAGA	CAGGTATTT	TATTTATACTTACAGC	CTGGTAAGT
	exon 3	U89	TTCTTTAAATTCTAGC	None	TCTTTTACATCCTAGC	None
U91	exon 1	U91	None	CAGGTTTGT	None	CTGGTTAGT
	exon 2	—	TATTTTTTCTTGTAGA	None	ATGGTTTGTTTTTAGA	None
U100	exon 1	U100	AAAATCTCTTCGCAGA	ACAGTAAGT	CGAAATTTTCACAAGA	ACGGTAAGG
	exon 2	U99	TTTAATTCTTCTAAGG	ATGGTAAGC	TTTAATTTATCGCAGC	ATGGTGAGT
	exon 3	U98	GTACCCGCTTATTAGT	AGTGTAAGT	ATTTATTCACTCAAGT	CGTGTAAGT
	exon 4	—	TATTTTTTTTTTTAGA	AATGTAAGA	TTCGTTTTTTGTCAGG	ACTGTAAGT
	exon 5	—	AATTGTGTTTCGCAGT	CAGGTAAAT	TTATGTTTCTAACAGA	ACGGTGAGT
	exon 6	—	GCTTCTTCATCCTAGA	TTGGTAATT	TCCGGTTATGCACAGC	ATGGTGAGC
	exon 7	—	TTTTTTTCATACCAGC	ACAGTGGAA	TTTCTTAATTTGCAGC	AGGGTGGGC
	exon 8	—	TTTTTTTAATTCTAGC	CATGTGAGT	TTTCGACCTGCCTAGA[a]	AATGTAAGT
	exon 9	—	ATTTCTCGTTCGCAGC	CAGGTGAGC	TTGATATTTGTTCAGT	TAGGTATTA
	exon 10	H8	CATTTTCTCTTTTAGT	None	TGTTTTTTTTTTAAGT	None

[a] Another acceptor site (GCTACCGCTTTTTAGC) is located farther upstream. It could extend coding similarity between HHV-7 and HHV-6 exon 8, but it is in an inappropriate reading frame.

coding regions extended upstream from the first initiation codon or did not extend to the stop codon, sequences were analysed for the presence of consensus donor and acceptor sites. In general, splicing was proposed when candidate sites were located similarly in the two genomes, so that two exons could in principle be joined in the same register and express the conserved, and not the nonconserved, polypeptide regions. This process was not taken to the same level of detail for the entire genome, but only for regions in which splicing had been demonstrated experimentally, for regions in which splicing was suspected, or for regions lacking obvious coding potential.

Table 4 summarizes splice sites proposed from the analysis, and lists JI ORFs forming exons in spliced genes. The evidence for splicing in U66, U7, and U15 is shown in Fig. 2. U66 is shown as an example of a gene that is almost certainly spliced at sites which have been predicted previously. It has counterparts in all herpesviruses in which the splice sites and coding potential are strongly conserved, and has been shown experimentally to be spliced in herpes simplex virus type 1 (Dolan *et al.,* 1991). Figure 2 shows that significant amino acid sequence similarity does not extend beyond the proposed splice sites into the intron, that the sites are in the same register in HHV-7 and HHV-6, and that the exons are joined appropriately.

Figure 2 also shows evidence for U7 and U15, neither of which was previously suspected of being spliced. Two substantial ORFs in HHV-6, U5, and U7 are represented in JI by a single, apparently fused ORF (U5/7). Observing the presence of US22 motifs in the U7 domain, Nicholas (1996) speculated that the fused protein might have a unique regulatory function in HHV-7. However, the data in Fig. 2 suggest that ORFs U7 and U5 form two exons of a single gene in both genomes (U7 in Fig. 1 and Table 2). Thus, in this scheme U7 is flanked at its 5′ end by U8 which, like exon 1, contains US22 motifs, and at its 3′ end

U66

HHV-7

H E T I K S I A L F A S C Y N I H v s n k f f l p t y y a f t d w t
ACGAAACAATAAAAAGTACAGCACTGTTTGCTAGTTGTTACAACACTCACgtaagtaacaaatttttcttacctacttactatgcgtttacagactggac

g k e y - - y y i k t q s l y r l k l r s f s s q S l R G Q S
aggaaaggagtattaa.....tagtattacataaaaactcagtctttatatcgtttaaaactccgttcatttttcttctcagAGTATACGTGGTCAGAG

HHV-6

H Q N I K S I A L F A S C Y N I H v s t i n l t l f s e i k n -
ATCAAAATATAAAGAGCACAGCACTTTTCGCCAGCTGCTACAATACACACgtaagtactatcaatttaacttttatttttcagaaataaaaaattaa.....

- k g r l s f h s s q n r l l s r l t p n s f p s q S l R G Q S
.....taaaaaggtcgtctttcgtttcacagttctcaaaatcgtcttttatcccgtttaacgccgaattcattcccctctcagAGTATACGCGGACAGAG

U7

HHV-7

E N L N R M L N G E S P I L R K K P R H M Y P R C D R y v k i k v
GAAAATTTGAACAGAATGTTAAATGGAGAGTCTCCGATTCTTCGGAAGAAACCACGGCATATGTATCCAAGGTGTGATAGgtatgtaaagatcaaagttc

l l f g i l y t i l l l m v p w m f f f r L L K N M P S l L F A U H
ttcttttcggtattctctacacaatactcttactgatggttccatggatgttcttttttcagATTATTGAAGAACATGCCAAGTATTCTGTTTGCAGTGCA

S S E I S N P L U Q S U I K F L H P I I I P N G D T E L K Y I U P
CTCATCTGAAATCAGTAATCCTCTTGTTCAAAGTGTTACAAAGTTTCTACATCCAATCATCATTCCAAACGGAGACACGGAACTGAAGTACATTGTTCCG

U T E S R L I N G L Q A S A A G R F G l K G L R L C S D G U I W N
GTGACAGAATCCAGACTAATCAACGGCCTTCAAGCATCAGCTGCTGGACGCTTTGGAATAAAAGGGCTAAGATTATGTTCAGATGGTGTTATTTGGAATA

R L I D Y E Y E M F K Y P S I F T R A D K F L L Q L R D L K F
GGTTGATAGACTATGAGTACGAAATGTTCAAGTATCCATCAACTTTCACAAGGGCTGATAAATTTCTTTTGCAGTTGCGCGATTTAAAATTT

HHV-6

- a s a y v r a k g q v g g c v l c
E N L N R M L N G E L P U L R S K P R H M C U R K D R w v d v f c
GAGAATCTCAATAGGATGTTGAATGGGGAGTTGCCCGTTCTCCGTAGTAAGCCTCGGCATATGTGCGTGCGAAAGGACAGgtgggtggatgtgttctgtg

r f p e r i l i r g w r l p a n f f f s y n r L U K D R S K l L F
v v f q r g -
tcgttttccagagaggataattattcgaggatggagactccccgctaacttttttttttcttacaacagGCTTGTGAAAGACCGTAGCAAAATTCTGTTT

A U R L D E E D S P T U K F I I K F L T P U F U G R L P A T N R F
GCGGTGCGTCTGGATGAAGAGGATTCACCGACGGTTAAATTTATCACAAAATTTCTCACGCCGGTATTCGTCGGTCGACTGCCCGCTACCAACAGGTTTG

U U P U S R A R L T N G L Q G T A A A R F G l K G L H P S S D C L U
TTGTTCCCGTCTCCCGCGCCAGGTTGACGAACGGTCTGCAGGGGACCGCAGCAGCTAGATTTGGCATTAAAGGACTACATCCCTCTTCGGACTGCTTGGT

W N I L U D Y E Y E T Y K Y P S I Y I R A D Q I A D M U K D L K F
GTGGAACATACTGGTGGATTATGAGTATGAAACTTACAAGTATCCTTCCACCTACATCAGAGCCGATCAGATTGCGGACATGGTGAAAGACTTAAAATTT

FIG. 2. Sequence-derived evidence for splicing in the coding regions of HHV-7 and HHV-6 genes U66, U7, and U15. Each part shows the relevant DNA sequence with appropriate conceptual translation products aligned from left to right. For U66, amino acid sequences extending to the stop codon defining the 3′ end of the upstream ORF and from the stop codon (hyphen) defining the 5′ end of the downstream ORF are shown. The DNA sequences begin and end at arbitrary points in the ORFs, and conserved amino acid residues are underlined. The proposed intron sequence (flanked by splice donor and acceptor sites as listed in Table 4) is shown in lower case in the DNA sequence, and amino acid residues excluded from the protein predicted to be expressed via splicing are also shown in lower case. The dots indicate a sizeable region of the intron that is not shown. For U7, the two ORFs are in the same frame in HHV-7 (and therefore are not divided by stop codons), but overlap in HHV-6. Double underlining indicates the initiation codon previously assigned to the downstream HHV-6 ORF (U5). The DNA sequences for U15 contain three exons and two introns, and extend from the proposed initiation codon in exon 1 to the stop codon in exon 3.

by U4, which is related to exon 2 (Table 2). The analysis also gave a somewhat weaker indication that U7 may be spliced at its 5′ end, as amino acid sequence similarity extends upstream from the previously proposed HHV-7 initiation codon. We were unable, however, to identify an upstream coding exon. U8 is a possibility, but this ar-

U15

HHV-7

```
M E T W R R Q R L Q E F R E L C P L Q I L M T L S N I I S K V E T
ATGGAAACCTGGAGAAGACAACGACTACAGGAATTTCGCGAGCTGTGTCCACTACAGATATTGATGACGTTATCTAATATTATATCTAAAGTGGAGACAA

I V I K V L F Q M D F N I T V R F I F S G L I L T T I U I K S V U I
TCTATATAAAATATCTTTTTCAAATGGATTTTAATACAACATATAGATTTATTTTTTCTGGATTAACATTAACCACAACTGTGACAAAGTCTGTAGTGAT

E A L F I I K R W Q E I K Q I F N L D V H K T E D C Y I U A Q F T
TGAAGCTTTGTTTATTATTAAGAGATGGCAAGAAATCAAGCAGATTTTCAATCTAGATGTCCACAAAACTGAGGATTGTTATATCGTCGCTCAATTCACA

H I P V K R -                                   - p r s f i f r K I
CACATACCTGTAAAACGgtgagtccacctctattatttctacacaaacgagtttcaaaattttattgtctaaccccgctcttttattttcagGAAAATCA

T A L L V M M I T K H E K Q L F L N M I V A v r k i i f k n n k c s
CAGCATTATTGTATATGATGACAACAAAGCATGAAAAGCAACTTTTCTTAAACATGATTTATGCGgtaagaaaaatcatttttaaaaatаataaatgttc

                                - F L E E S H L R L G D D E H E N A I M F F
s v k d l l t f f f f s f -
ttctgtaaaagatcttttaacatttttttttctttagTTTCTAGAAGAAAGTCATCTGAGGCTTGGAGACGATGAGCATGAAAACGCGATAATGTTTTTTT

S Y I E R L Q L I R D V L I E I I Q K L K N V E I N Q I I A L V L S
CATACATAGAACGACTCCAACTGACCAGAGATGTTTTAATTGAGATTATTCAAAAGCTAAAAAACGTGGAAATTAATCAAACAATTGCCCTTGTGTTGTC

V N E L A K -
ATACAATGAATTAGCTAAATAA
```

HHV-6

```
M D V W K R Q R L Q E C R E L C P L P V L M S L S N M F S K I E I
ATGGATGTGTGGAAGCGTCAACGGCTTCAAGAATGCCGTGAATTGTGTCCTTTGCCTGTATTAATGTCACTGTCGAATATGTTTTCAAAAATCGAAATCG

V Y V K Y L F K M D F S I M Y R Y I L P A L I L S M I U I K S L U I
TATACGTTAAATATCTATTTAAAATGGACTTTTCTACTATGTATAGATATATTTTACCGGCTCTTACGTTGAGTATGACGGTTACAAAATCCCTAGTTAT

E M L F I L K R W E D I D Q F F R L N I R K V N D C F I U A Q F N
TGAAATGTTATTTATTTTAAAAAGATGGGAAGATATTGATCAATTTTTTAGATTGAACATCCGGAAAGTAAACGACTGTTTCATCGTAGCTCAGTTCAAC

H I P I K R w v I i -                           - c m p i n v f f f y r K L I
CATATTCCTATAAAAAGgtgggtactgatttaagtatacgtggaagcgagcgggctaatgtatgccgattaatgtgtttttctttttacagGAAGTTAATT

V L L V M L I S R Q E K Q L F L N M I V A v s I i I y s r f q d i
GTGTTGTTGTATATGTTAACCAGTAGACAAGAAAAACAGCTTTTCCTCAATATGATATATGCGgtgagtttgattttgtacagtagatttcaagatatat

I n c n -
taaattgtaattaggtgcttggtccattaggttgctagactatacttttctggtggtaatttagtgtttaatgtaaagttatattggattggatggataa

                                - F L E K S H L R L G D D E E Q N A I R E F S V V
aattgatggtttgtgtttttttttttagTTTTTAGAGAAGAGTCACCTTAGACTTGGTGATGATGAAGAACAGAATGCTATTCGTTTTTTTTCTTACGTT

D D L H L I R D I L L E M I H K L K N T E I N Q I M E L L L S V N
GATGATTTACACCTAACGCGGGATATTTTATTGGAGATGATCCACAAGCTCAAGAATACGGAAATCAATCAAACAATGGAACTTTTATTATCGTACAATG

E L A R -
AATTGGCTAGATAA
```

FIG. 2—*Continued*

rangement would result in a protein with an unusual structure (i.e., two contiguous US22 domains). Figure 2 also strongly supports the presence of three exons in U15. Amino acid sequence conservation is high, and splice sites are located identically in the two genomes.

Splicing events were previously suspected of occurring in five genes for which experimental data were not available: DR1, DR6, U12, U17, and U79. It had been reported for HHV-6 and JI that DR1 is related to DR6 and DR2 to DR7, and that each ORF contains motifs characteristic of the US22 gene family. Nicholas (1996) noted that JI DR1 and DR2 lack suitable initiation codons, in contrast to their HHV-6 counterparts, speculated that these might be supplied by splicing, and commented that splicing of DR1 to DR2 and DR6 to DR7 would result in proteins containing the US22 motifs in their usual order. Our analysis supports the view that DR1 and DR2 form exons in a spliced gene (DR1 in Table 2) and that DR1 may be spliced at its 5′ end to a noncoding exon, but indicates that DR1 supplies its own initiation codon. Similarly, we agree that DR6 and DR7 may be spliced (DR6 in Table 2), and found no evidence for splicing at

the 5′ end of DR6. The donor and acceptor sites proposed for HHV-6 U12 in Table 4 differ from those of Gompels *et al.* (1995); splicing was not proposed for the JI counterpart by Nicholas (1996). With regard to U17, our analysis supports the previous assignments of Nicholas and Martin (1994) and Nicholas (1996).

Schiewe *et al.* (1994) derived detailed splicing patterns for HHV-6 U90 and U91 from transcript mapping and cDNA cloning experiments. Immediate early gene U90 contains three exons, and gene U91 contains two. The HHV-7 and HHV-6 U91 proteins contain two pronounced hydrophobic regions (one near the N terminus and the other about two-thirds through the protein) and therefore are likely to be associated with membranes. Both have a single consensus site for N-linked glycosylation in the region between the hydrophobic domains, and the HHV-7 protein has two additional sites near the C terminus. Evidence that U91 encodes putative membrane glycoproteins in HHV-7 and HHV-6 has not been reported previously.

The region at the right end of U contains several short ORFs which have been shown by sequencing a cDNA clone to be expressed in HHV-6 as a multiply spliced mRNA encoding the virion envelope glycoprotein gp105 (Pfeiffer *et al.,* 1995). The mRNA contains ten coding exons (here termed exons 1–10), five of which were identified as ORFs in the HHV-6 sequence, plus two upstream noncoding exons. Nicholas (1996) described four ORFs in this part of the JI genome, but was unable to deduce a splicing pattern. Our attempt to decipher the arrangement of the RK gene is given in Table 2. There is less certainty at some points because polypeptides encoded by exons 1 and 10 are poorly conserved. It was also necessary to propose that the exon 8 and 10 acceptor sites are not similarly located in HHV-7 and HHV-6, although they are in the same register in each genome. Indeed, it is curious that use of an alternative acceptor site for HHV-6 exon 8 located upstream from the site mapped by Pfeiffer *et al.* (1995) could significantly extend the similarity with the HHV-7 exon 8 polypeptide; details are given in Table 4. The site is in an incorrect reading frame, however, so this point remains unresolved. The status of upstream noncoding exons is unknown.

Various other possible splice sites were noted during the analysis, including those supporting splicing of U12 exon 1 to U13 and of the 3′ end of U19 to an unidentified downstream exon. The evidence in these instances was weak, however, and the sites must be considered much more speculative than those listed in Table 4. In a few regions where splicing might reasonably be expected, such as that containing U24 and U24a, no evidence was forthcoming.

All spliced genes proposed above except U66 are noncore genes, although some have counterparts in HCMV and MCMV. Indeed, it was the evidence for splicing in certain HCMV genes that led to initial suspicions of splicing in the HHV-7 and HHV-6 counterparts (e.g., U17). Some of the putatively spliced genes in HHV-7 and HHV-6 have counterparts in HCMV and MCMV that are not known to be spliced (e.g., U7) or are spliced differently (e.g., U79). The evolutionary distances involved, however, preclude easy identification of corresponding splicing patterns in the HCMV and MCMV genes. Also, database searches conducted using proteins encoded by putative spliced mRNAs did not reveal significant similarities to proteins other than those reported previously.

The degree of confidence in the splicing patterns proposed above rests on the level of amino acid sequence conservation, the presence of potential splice sites in corresponding locations in both genomes, and, where available, on experimental corroboration. In summary, we consider that each of the proposed HHV-7 splice sites in Table 4 is a strong candidate worthy of experimental investigation. Nevertheless, complex splicing patterns, use of alternative splice sites, long-range splicing, and involvement of noncoding exons are very likely to have escaped detection. Consequently, we speculate that splicing is probably more common in HHV-7 and HHV-6 than is suggested from our analysis.

MATERIALS AND METHODS

Preparation and purification of DNA

Human cord blood mononuclear cells were treated with 10 μg/ml phytohemagglutinin for 2 days in RPMI 1640 medium containing 10% (v/v) fetal calf serum, 50 μg/ml gentamycin (Frenkel *et al.,* 1990; Berneman *et al.,* 1992b; Black and Pellett, 1993). The cells were then incubated for 2 h at 37°C with concentrated aliquots of HHV-7(RK). Following adsorption, the cells were diluted with medium to 10^6 cells/ml (approximately 300 ml total), and incubated further at 37° (Frenkel and Rapaport, 1995). The typical CPE of HHV-7 infection, characterized by ballooning of cells and the appearance of limited syncytia, became evident in the culture. At 14 days after infection, cells were harvested by centrifugation, rinsed in phosphate-buffered saline and pelleted once more. The cells were resuspended in 10 mM Tris–HCl, pH 7.5, 10 mM NaCl, 15 mM $MgCl_2$, 0.6% (v/v) Nonidet P40, Dounce homogenized, and separated into nuclear and cytoplasmic fractions by centrifugation. The fractions were treated with 0.5% (w/v) sodium deoxycholate and incubated with 50 μg/ml DNase I and 10 μg/ml RNase A, and capsids were prepared by density centrifugation on sucrose gradients as described by Gibson and Roizman (1972) and Vlazny *et al.* (1982). DNA was extracted from capsids as described by Di Luca *et al.* (1990).

In order to confirm the presence of HHV-7 genomic DNA and to check for contamination with smaller fragments, DNA preparations were assessed by agarose gel

electrophoresis of untreated material and material digested by restriction endonucleases. Genomic DNA was precipitated from one sample obtained from the cytoplasmic fraction by incubating with 8% (w/v) polyethylene glycol 6000, 1 M NaCl for 16 h at 0°C. The DNA was pelleted by centrifugation for 5 min at 11,000*g* in a microcentrifuge, washed twice with 70% ethanol, air dried, and resuspended in a small volume of 10 mM Tris–HCl, pH 7.5, 1 mM EDTA.

DNA sequencing and sequence analysis

Random DNA fragments were generated by sonication of purified genomic DNA. Fragments 400–700 bp in size were excised from an agarose gel, purified using a Geneclean II kit (Bio 101, Inc.), treated with T4 DNA polymerase in the presence of the four deoxynucleoside triphosphates, and cloned into the *Sma*I site of bacteriophage M13mp19. DNA templates were prepared from recombinant plaques produced by transfection of ligated DNA into *Escherichia coli* DH5αF′ (Life Technologies Ltd.), and sequenced by the dideoxynucleotide chain termination technology (Sanger *et al.,* 1977) as described by Davison (1991). Autoradiographs were read using a Summagraphics digitizer and the database was compiled using Staden's sequence analysis program (Staden, 1987). The sequence was edited by reference to autoradiographs and analysed using programs from the Wisconsin Package Version 9.0, Genetics Computer Group (Madison, WI). The sequence has been deposited with the GenBank Data Library under Accession No. AF037218.

Sequence comparisons were carried out using the sequences of HHV-7(RK), HHV-7(JI) (Nicholas, 1996; Accession No. U43400, Version 5), and HHV-6A(U1102) (Gompels *et al.,* 1995; Accession No. X83413, Version 16). For certain comparisons, published sequences from other HHV-6A strains were also utilized.

ACKNOWLEDGMENTS

We are grateful to Duncan McGeoch for criticizing the manuscript. We thank the Israel Academy of Science, the Israel Ministry of Health, and the Wolfson Foundation for their support of this research.

REFERENCES

Berneman, Z. N., Ablashi, D. V., Li, G., Eger-Fletcher, M., Reitz, M. S., Jr., Hung, C.-L., Brus, I., Komaroff, A. L., and Gallo, R. C. (1992a). Human herpesvirus 7 is a T-lymphotropic virus and is related to, but significantly different from, human herpesvirus 6 and human cytomegalovirus. *Proc. Natl. Acad. Sci. USA* 89, 10552–10556.

Berneman, Z. N., Gallo, R. C., Ablashi, D. V., Frenkel, N., Katsafanas, G., Kramarsky, B., and Brus, I. (1992b). Human herpesvirus 7 (HHV-7) strain JI: Independent confirmation of HHV-7. *J. Infect. Dis.* 166, 690–691.

Black, J. B., and Pellett, P. E. (1993). Human herpesvirus 7. *Rev. Med. Virol.* 3, 217–223.

Chee, M. S., Bankier, A. T., Beck, S., Bohni, R., Brown, C. M., Cerny, R., Horsnell, T., Hutchison, C. A., III, Kouzarides, T., Martignetti, J. A., Preddie, E., Satchwell, S. C., Tomlinson, P., Weston, K. M., and Barrell, B. G. (1990). Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. *Curr. Top. Microbiol. Immunol.* 154, 125–169.

Davison, A. J. (1991). Experience in shotgun sequencing a 134 kilobase pair DNA molecule. *DNA Sequence* 1, 389–394.

Di Luca, D., Katsafanas, G., Schirmer, E. C., Balachandran, N., and Frenkel, N. (1990). The replication of viral and cellular DNA in human herpesvirus 6-infected cells. *Virology* 175, 199–210.

Dolan, A., Arbuckle, M., and McGeoch, D. J. (1991). Sequence analysis of the splice junction in the transcript of herpes simplex virus type 1 gene UL15. *Virus Res.* 20, 97–104.

Frenkel, N., Schirmer, E. C., Wyatt, L. S., Katsafanas, G., Roffman, E., Danovich, R. M., and June, C. H. (1990). Isolation of a new herpesvirus from human $CD4^+$ T cells. *Proc. Natl. Acad. Sci. USA* 87, 748–752.

Frenkel, N., and Rapaport, D. (1995). Detection of human herpesviruses 6 and 7 by PCR. *In* "PCR: Protocols for Diagnosis of Human and Animal Virus Diseases" (Y. Becker and G. Davis, Eds.), pp. 243–251. Springer, Berlin.

Frenkel, N., and Roffman, E. (1996). Human herpesvirus 7. *In* "Fields Virology" (B.N. Fields, D.M. Knipe, and P.M. Howley, Eds.), 3rd ed., pp. 2609–2622. Lippincott-Raven, Philadelphia.

Gibson, W., and Roizman, B. (1972). Proteins specified by herpes simplex virus. VIII. Characterization and composition of multiple capsid forms of subtypes 1 and 2. *J. Virol.* 10, 1044–1052.

Gompels, U. A., Nicholas, J., Lawrence, G., Jones, M., Thomson, B. J., Martin, M. E. D., Efstathiou, S., Craxton, M., and Macaulay, H. A. (1995). The DNA sequence of human herpesvirus-6: Structure, coding content, and genome evolution. *Virology* 209, 29–51.

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

McGeoch, D. J., and Cook, S. (1994). Molecular phylogeny of the Alphaherpesvirinae subfamily and a proposed evolutionary timescale. *J. Mol. Biol.* 238, 9–22.

Nicholas, J. (1994). Nucleotide sequence analysis of a 21-kbp region of the genome of human herpesvirus-6 containing homologues of human cytomegalovirus major immediate-early and replication genes. *Virology* 204, 738–750.

Nicholas, J. (1996). Determination and analysis of the complete nucleotide sequence of human herpesvirus 7. *J. Virol.* 70, 5975–5989.

Nicholas, J., and Martin, M. E. D. (1994). Nucleotide sequence analysis of a 38.5-kilobase-pair region of the genome of human herpesvirus 6 encoding human cytomegalovirus immediate-early gene homologs and transactivating functions. *J. Virol.* 68, 597–610.

Okuno, T., Takahashi, K., Balachandra, K., Shiraki, K., Yamanishi, K., Takahashi, M., and Baba, K. (1989). Seroepidemiology of human herpesvirus 6 infection in normal children and adults. *J. Clin. Microbiol.* 27, 651–653.

Pfeiffer, B., Thomson, B., and Chandran, B. (1995). Identification and characterization of a cDNA derived from multiple splicing that encodes envelope glycoprotein gp105 of human herpesvirus 6. *J. Virol.* 69, 3490–3500.

Rawlinson, W. D., Farrell, H. E., and Barrell, B. G. (1996). Analysis of the complete DNA sequence of murine cytomegalovirus. *J. Virol.* 70, 8833–8849.

Roizman, B., Carmichael, L. E., Deinhardt, F., de-The, G., Nahmias, A. J., Plowright, W., Rapp, F., Sheldrick, P., Takahashi, M., and Wolf, K. (1981). Herpesviridae: Definition, provisional nomenclature, and taxonomy. *Intervirology* 16, 201–217.

Ruvulo, V. R., Berneman, Z., Seccheiro, P., and Nicholas, J. (1996). Cloning, restriction endonuclease mapping and partial sequence analysis of the genome of human herpesvirus 7 strain JI. *J. Gen. Virol.* 77, 1901–1912.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with

chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.

Schiewe, U., Neipel, F., Schreiner, D., and Fleckenstein, B. (1994). Structure and transcription of an immediate-early region in the human herpesvirus 6 genome. *J. Virol.* 68, 2978–2985.

Secchiero, P., Nicholas, J., Deng, H., Xiaopeng, T., van Loon, N., Ruvulo, V. R., Berneman, Z. N., Reitz, M. S. Jr., and Dewhurst, S. (1995). Identification of human telomeric repeat motifs at the genome termini of human herpesvirus 7: Structural analysis and heterogeneity. *J. Virol.* 69, 8041–8045.

Singer, O., and Frenkel, N. (1997). Human herpesvirus 7 (HHV-7) DNA: Analyses of clones spanning the entire genome. *Arch. Virol.* 142, 287–303.

Staden, R. (1987). Computer handling of DNA sequencing projects. *In* "Nucleic Acid and Protein Sequence Analysis: A Practical Approach" (M. J. Bishop, and C. J. Rawlings, Eds.), pp. 173–217. IRL Press, Oxford.

Tanaka, K., Kondo, T., Torigoe, S., Okada, S., Mukai, T., and Yamanishi, K. (1994). Human herpesvirus 7: Another causal agent for roseola (exanthem subitum). *J. Pediatr.* 125, 1–5.

van Loon, N., Dykes, C., Deng, H., Dominguez, G., Nicholas, J., and Dewhurst, S. (1997). Identification and analysis of a lytic-phase origin of DNA replication in human herpesvirus 7. *J. Virol.* 71, 3279–3284.

Vlazny, D. A., Kwong, A., and Frenkel, N. (1982). Site-specific cleavage/packaging of herpes simplex virus DNA and the selective maturation of nucleocapsids containing full-length viral DNA. *Proc. Natl. Acad. Sci. USA* 79, 1423–1427.

Wyatt, L. S., Rodriguez, W. J., Balachandran, N., and Frenkel, N. (1991). Human herpesvirus 7: Antigenic properties and prevalence in children and adults. *J. Virol.* 65, 6260–6265.

Yamanishi, K., Okuno, T., Shiraki, K., Takahashi, M., Kondo, T., Asano, Y., and Kurata, T. (1988). Identification of human herpesvirus-6 as a causal agent for exanthem subitum. *Lancet* i, 1065–1067.