JOURNAL OF VIROLOGY, May 2001, p. 4854–4870 Vol. 75, No. 10
0022-538X/01/$04.00+0 DOI: 10.1128/JVI.75.10.4854–4870.2001

Analysis and Characterization of the Complete Genome of Tupaia (Tree Shrew) Herpesvirus

UDO BAHR AND GHOLAMREZA DARAI*

Institut für Medizinische Virologie, Universität Heidelberg, D-69120 Heidelberg, Germany

Received 13 November 2000/Accepted 26 February 2001

The tupaia herpesvirus (THV) was isolated from spontaneously degenerating tissue cultures of malignant lymphoma, lung, and spleen cell cultures of tree shrews (*Tupaia* spp.). The determination of the complete nucleotide sequence of the THV strain 2 genome resulted in a 195,857-bp-long, linear DNA molecule with a G+C content of 66.5%. The terminal regions of the THV genome and the loci of conserved viral genes were found to be G+C richer. Furthermore, no large repetitive DNA sequences could be identified. This is in agreement with the previous classification of THV as the prototype species of herpesvirus genome group F. The search for potential coding regions resulted in the identification of 158 open reading frames (ORFs) regularly distributed on both DNA strands. Seventy-six out of the 158 ORFs code for proteins that are significantly homologous to known herpesvirus proteins. The highest homologies found were to primate and rodent cytomegaloviruses. Biological properties, protein homologies, the arrangement of conserved viral genes, and phylogenetic analysis revealed that THV is a member of the subfamily *Betaherpesvirinae*. The evolutionary lineages of THV and the cytomegaloviruses seem to have branched off from a common ancestor. In addition, it was found that the arrangements of conserved genes of THV and murine cytomegalovirus strain Smith, both of which are not able to form genomic isomers, are colinear with two different human cytomegalovirus (HCMV) strain AD169 genomic isomers that differ from each other in the orientation of the long unique region. The biological properties and the high degree of relatedness of THV to the mammalian cytomegaloviruses allow the consideration of THV as a model system for investigation of HCMV pathogenicity.

The family *Herpesviridae* comprises more than 100 different virus species with a worldwide occurrence in all taxonomic groups of vertebrates. The supposed roots of this virus family are in very early evolutionary times, and a long period of development has resulted in the appearance of extremely well host-adapted virus species (47, 61, 62, 63), often more than one in a single host, for example, the eight different human herpesviruses HSV-1 (herpes simplex virus type 1), HSV-2, varizella-zoster virus, HCMV (human cytomegalovirus), EBV (Epstein-Barr virus), HHV-6 (human herpesvirus 6), HHV-7, and HHV-8, that are adapted to different cellular and molecular niches in the same host species (88).

A member of the large herpesvirus family is the *Tupaia herpesvirus* (THV) that infects tree shrews (*Tupaia* spp., family *Tupaiidae*), a group of primitive higher mammals (*Proteutheria, Scandentia*) that is supposed to have diverged at the base of the primate evolutionary tree (52, 69). Tree shrews were originally distributed in Southeast Asia and are used worldwide as laboratory animals in neurological and physiological research. THV was isolated by Mirkovic et al. in 1970 (68) from a spontaneously degenerating lung tissue culture of a tree shrew and subsequently classified as a herpesvirus by electron microscopic examination (57). From 1977 to 1985, six additional isolates were isolated from malignant lymphoma tissue cultures and degenerating spleen cell cultures of tree shrews (19, 20, 21, 22, 24, 49). The seven THV isolates were grouped into five strains (THV strains 1 to 5) according to their restriction endonuclease cleavage patterns. Molecular cloning and physical mapping of the genome of THV strain 2 was performed, and a complete genome library was established (49). THV strain 2 was isolated in 1979 from a spontaneously degenerating malignant lymphoma cell culture by Darai et al. (19).

THV particles show the classical morphology of herpesviruses (19, 61, 73) and contain a linear double-stranded DNA genome of about 200 kbp (21, 49). The detection of concatemeric viral DNA molecules in infected cells (50) corresponds to the rolling-circle model of herpesvirus genome replication. The herpesviruses have been classified into six genome groups (A to F) (72) according to the presence and arrangement of large repetitive DNA sequences. THV is the prototype species of genome group F, which is characterized by a unique DNA sequence without any extended repetitive DNA elements. The family *Herpesviridae* is subdivided into the subfamilies *Alpha-*, *Beta-*, and *Gammaherpesvirinae* according to the length of the replication cycle, speed of spreading in cell culture, host range, and location of latency, which is an important biological characteristic of herpesviruses (73). According to its biological properties (19, 20, 22, 24) and in agreement with recent data (5, 82), THV is supposed to be a member of the subfamily *Betaherpesvirinae*.

THV infections cause a remarkable variety of different clinical pictures in tree shrews, ranging from inapparent to deadly infections and the development of malignant lymphomas (19, 23, 41). Based on the evolutionary stage of tree shrews and the biological and genomic properties of THV, the elucidation of the viral coding strategy is of particular interest. The characterization of the primary structure of the whole genome of THV strain 2 isolated from a malignant lymphoma (19) is the subject of this report. This study allowed the determination of

* Corresponding author. Mailing address: Institut für Medizinische Virologie, Universität Heidelberg, Im Neuenheimer Feld 324, D-69120 Heidelberg, Germany. Phone: 49-6221-56-5008. Fax: 49-6221-56-4104. E-mail: g.darai@urz.uni-heidelberg.de.

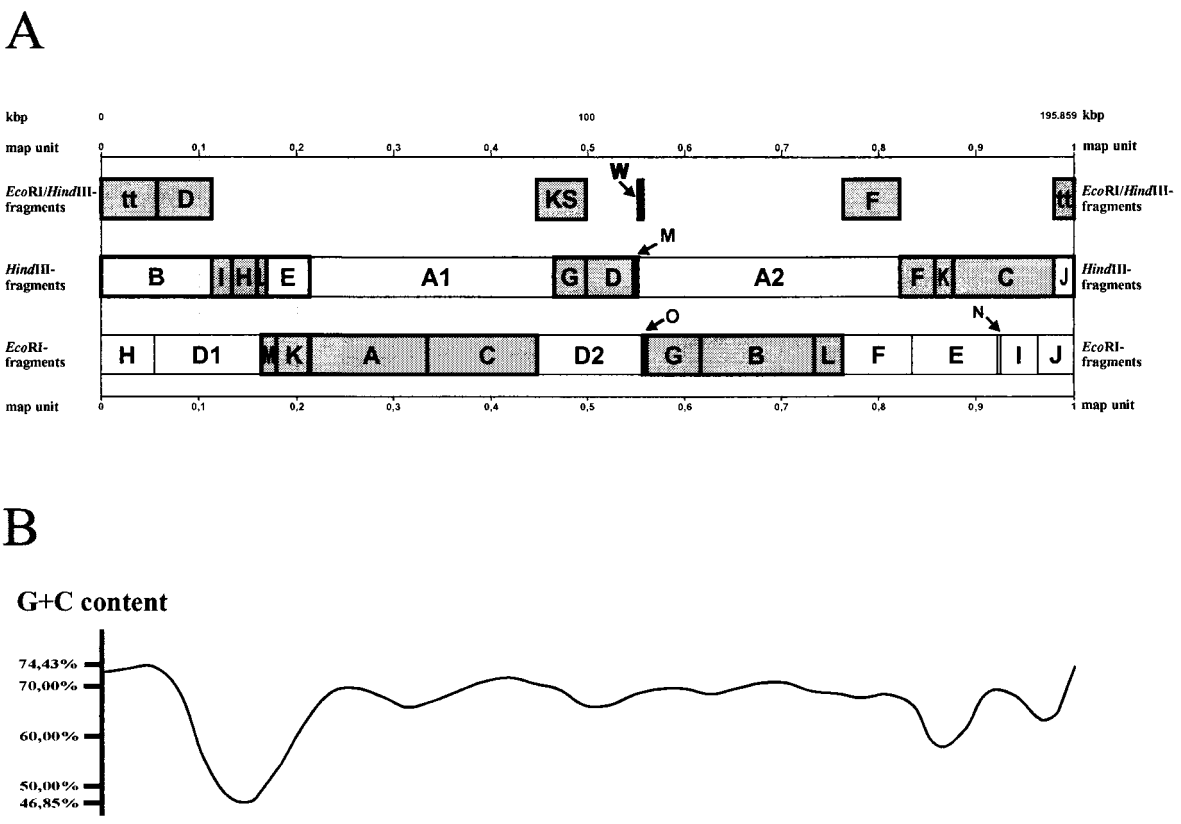


FIG. 1. (A) Physical map of the THV strain 2 genome consisting of *Eco*RI, *Hin*dIII, and *Eco*RI/*Hin*dIII DNA fragments. The THV DNA fragments that are used to determine the nucleotide sequence of the whole THV genome are shaded and bordered in bold. Beneath the physical map, the G+C content over the course of the whole THV genome is shown (B). It varies between 46.85 and 74.43%.

the final phylogenetic placement of THV within the family *Herpesviridae*.

MATERIALS AND METHODS

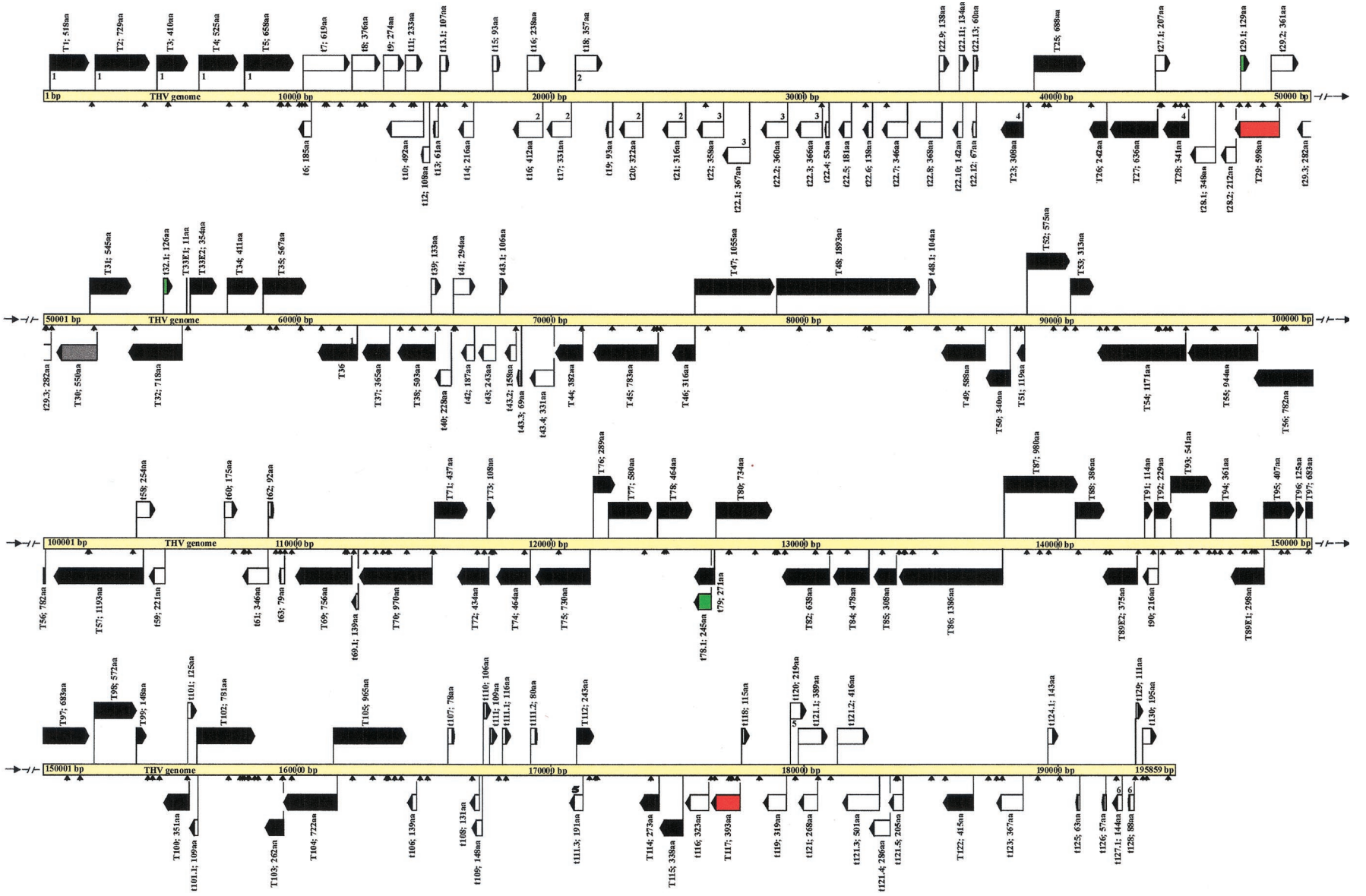
Viral DNA and genomic library. Propagation of THV on tupaia baby fibroblasts and isolation of viral DNA were carried out as described previously (22). Recombinant plasmids harboring specific DNA sequences of the THV genome were obtained from a defined genome library as described elsewhere (49).

Strategy of determination of the THV genome nucleotide sequence. Determination of the complete nucleotide sequence of the THV genome was accomplished by analysis of the DNA nucleotide sequences of recombinant plasmids harboring specific *Eco*RI, *Hin*dIII, and *Eco*RI/*Hin*dIII fragments of the THV genome (Fig. 1) that form the entirety of the THV genome. The nucleotide sequences of the individual recombinant plasmids were determined by primer walking (86). The correctness of the physical map shown in Fig. 1 was confirmed by amplification and sequencing of the genome regions of original THV DNA around the endonuclease restriction sites, which allowed the assembly of the nucleotide sequences of the individual THV fragments resulting in the nucleotide sequence of the whole THV genome.

Enzymes and DNA isolation. The restriction endonucleases were purchased from Roche Diagnostics GmbH (Mannheim, Germany). Incubations were carried out according to standard procedures for each enzyme. The recombinant plasmids harboring the DNA sequences of the *Eco*RI, *Hin*dIII, and *Eco*RI/*Hin*dIII fragments that were used to determine the complete nucleotide sequence of the THV genome were purified using the Qiagen Plasmid Midi Prep (Qiagen GmbH, Hilden, Germany) procedure. The PCR products were purified using Micro-Spin S-300 HR columns (Pharmacia Biotech).

DNA sequencing. The recombinant plasmids and the purified PCR products were sequenced using the DyeDeoxy Terminator *Taq* cycle sequencing technique (Ready Reaction DyeDeoxy Terminator Cycle Sequencing Kit; Applied Biosystems GmbH, Weiterstadt, Germany) and a 373A "Extended" DNA sequencer (Applied Biosystems) as described previously (86). The nucleotide sequence of the THV genome was determined by primer walking. The Sequence Navigator software (Version 1.01; Applied Biosystems) was used to assemble the nucleotide sequences obtained from individual sequencing reactions.

Computer-assisted analysis. The strategies used to identify THV genes likely to encode were based on those used in the sequence analysis of other herpesviruses (71). The major criterion for identifying a coding sequence was the presence of an open reading frame (ORF) with a minimum length of 300 bp and less than 60% overlap with adjacent ORFs. The presence of ORFs with more than 60% overlap in Fig. 2 has its explanation in similar lengths and the absence of convincing identification criteria which distinguish between the corresponding ORFs. In addition, analysis of codon usage, the presence of consensus promoter sequences, and homology to known genes were used to support initial ORF selection. Gaps between ORFs were inspected for smaller ORFs and were included in the final ORF map (Fig. 2) if they satisfied the other criteria used for ORF selection. The identification of ORFs was performed with the program TRANSLATE of the the PC/GENE program, release 6.85 (Intelligenetics Inc., Mountain View, Calif.); the decisive criterion for the selection of an ORF was the presence of an ATG start codon and a stop codon (TAA, TGA, TAG) at the end. Searches of potential THV proteins for homology to known proteins were performed by applying BLAST (basic local alignment search tool) (3) and FSTPSCAN of PC/GENE to SWISSPROT database release 39. Protein alignments were carried out with the CLUSTAL program (39). Examination of the DNA sequence for transcription signals was performed by using the search features of the program EUKPROM of PC/GENE 6.85. Protein motif searches were performed with the program PROSITE of PC/GENE 6.85 and PROSITE release 16.25 on the ExPASy Molecular Biology Server (10, 40). The sequence was analyzed for tandem and inverted repeats by using the program REPEATS of PC/GENE 6.85. The phylogenetic trees were calculated using the personal computer programs ClustalW (version 1.64b) (85), Kitsch (version 10.0), Protdist (version 10.0), and Treeview (version 1.5.2).


1 bp
THV genome
10000 bp
20000 bp
30000 bp
40000 bp
50000 bp
50001 bp
60000 bp
70000 bp
80000 bp
90000 bp
100000 bp
100001 bp
110000 bp
120000 bp
130000 bp
140000 bp
150000 bp
150001 bp
160000 bp
170000 bp
180000 bp
190000 bp
195859 bp
No homolgy detected
Homolgy to HCMVA and MCMVS protein
Only homolgy to HCMVA protein
Only homolgy to MCMVS protein
Highest homolgy to HHV6 protein
Hairpin loop (Stem >7bp; Loop: 3-20 bases)

Nomenclature. All of the ORFs that are likely to be coding sequences are listed in Table 1. They are numbered in the order of their appearance when the genome sequence is analyzed from base pair 1 to base pair 195,859. ORFs that are oriented to the right are marked with the letter R, and ORFs that are oriented to the left are marked with the letter L. In addition, ORFs are marked with the letter T for THV and numbered according to homologous proteins of HCMV strain AD169 (HCMVA) and murine cytomegalovirus strain Smith (MCMVS). THV ORFs that are assumed to be coding sequences but have no homology to known proteins are marked with the letter t and labeled so as to fill the gaps between the homologous ORFs.

RT PCR. Tupaia baby fibroblasts were infected with THV strain 2. Total cellular RNA was isolated 14 h postinfection. RNA isolation was performed using the guanidinium-cesium chloride method as described previously (74). The reverse transcription step was carried out using the RNA LA PCR Kit, version 1.1 (Takara Shuzo Co., Shiga, Japan). PCR was performed using 0.5 fmol of the template DNA in 100-μl volumes containing 1.5 mM $MgCl_2$, 12.5 nmol of each deoxynucleoside triphosphate, 50 pmol of each primer, and 2.5 U of *ExTaq* DNA polymerase (Takara Shuzo Co.). A total of 35 cycles were run in an automated temperature cycling reactor (Genius; Techne, Cambridge, United Kingdom) under cycling conditions of 96°C for 30 s, 60°C for 1 min, and 72°C for 2 min per cycle. Specific oligonucleotide primers were designed in order to amplify the coding region around the splice sites of THV genes T33 and T89. The following primers were used in the reverse transcriptase PCR (RT-PCR) experiments: 5′-CCATGGACGTCCTGCTGGCTC-3′ (primer 1) and 5′-CCCACGGTGCA GCTGGTGTAG-3′ (primer 2) for T33 and 5′-AAGCACGTTTCCCAGTTCG TCC-3′ (primer 3) and 5′-GAGTTTGGTCAGGAAGCAGGTG-3′ (primer 4) for T89. Primers 2 and 4 were used for first-strand synthesis of the cDNA by RT reaction.

Nucleotide sequence accession number. The complete DNA sequence determined in this study has been submitted to the GenBank database and assigned accession number AF281817.

RESULTS

Features of the complete nucleotide sequence of the THV genome. The nucleotide sequences of the accentuated THV fragments in Fig. 1 were determined by automated cycle sequencing and primer walking (86). The assembly of the nucleotide sequences of all of the THV fragments used according to the physical map of the viral genome (Fig. 1) resulted in determination of the complete DNA nucleotide sequence of the THV genome, comprising 195,859 bp. The DNA sequences of the genomic termini were defined by Albrecht et al. in 1985 (2) and were used to determine the left and the right ends of the genome in this study. Altogether, 1,473 sequencing reactions with a total of 684,259 determined bases were performed to obtain the nucleotide sequence of the complete THV genome. Both DNA strands were sequenced independently, and each nucleotide was determined with an average redundancy of 1.75. The average G+C content of the whole THV DNA molecule was found to be 66.5%. As shown in Fig. 1, the G+C distribution is not constant in the viral genome (46.85 to 74.43%). Analysis of the codon usage of THV revealed that the third base of codons of potential viral genes is almost exclusively a G or C.

Repetitive elements of the THV genome. About 2,000 direct and 1,611 inverted repeats were identified. However, all of these repeats were no longer than 25 bp and occurred in short tandem arrangements. Some of the inverted repeats are supposed to form stable hairpin structures and are drawn in the ORF map of THV in Fig. 2. Furthermore, the analysis of the THV nucleotide sequence resulted in the identification of rare direct repeats between 25 and 105 bp in length. They are located between nucleotide positions 64165 and 64334, 108477 and 109442, 140381 and 140452, 186262 and 186305, and 193606 and 194265. The repetitive DNA sequences of the THV genome are restricted and are not comparable to the large repetitive elements of other herpesvirus genomes that are classified into genome groups A to E. This is in agreement with the previous nomination of THV as the prototype species of herpesvirus genome group F (49, 72).

ORFs of the THV genome. THV genome analysis revealed 582 ORFs with a possible coding capacity of more than 40 amino acids. The distribution of the ORFs on the two DNA strands is very regular, with 286 oriented to the left and 296 oriented to the right. Altogether, 158 ORFs were selected to be actual coding sequences according to the criteria used to determine herpesvirus genes (see Materials and Methods) and are listed in Table 1. Potential sites and signatures of the hypothetical viral proteins and the highest homologies to known proteins are also given in Table 1. The 158 ORFs selected to encode viral proteins are depicted in the ORF map of the THV genome in Fig. 2. Those viral gene products that are homologous to known herpesvirus proteins are shown as black or colored arrows. It is evident that the homologous, conserved ORFs are accumulated in the center of the genome. t6-t22.13, t39-t43.4, t58-t63, and t106-t130 are genome regions with ORFs that encode THV-specific proteins with no known homologues. RT-PCR experiments revealed that T33 and T89 consist of two exons. Comparative amino acid analysis to known proteins showed that T36 could be spliced, but it was not possible to prove this by RT-PCR experiments.

Conserved THV proteins show the highest homologies to known proteins of HCMVA (13) and MCMVS (71) as representatives of the evolutionary lineages of primate and rodent cytomegaloviruses, respectively. The homology values of the THV proteins to the proteins of these two virus species and the pertinent potential functions are summarized in Table 1. The proteins with the highest homologies are underlined, and it is clear that they are distributed equally between HCMVA and MCMVS.

All of the homologous THV ORFs are classified into functional groups and plotted according to their highest values of identity to known herpesvirus proteins (Fig. 3). The number of proteins gets smaller when the identity values rise, with T89 (viral terminase) being the only protein with 60 to 70% identity to known herpesvirus proteins. None of the functional protein groups show an extraordinary distribution of identities or strikingly low or high conservation. As a rule, they are almost constantly distributed over the range of identity values, with a decreasing number of proteins when the values get higher.

Comparison of the THV ORFs that are supposed to be

FIG. 2. Coding strategy of the THV genome. ORFs that code for proteins with homologies to known HCMVA and MCMVS proteins are depicted as black arrows, and those with no detected homologies are shown as white arrows. ORFs that show the highest homology only to an HCMVA or an MCMVS protein are in red or green, respectively. The orientation of each arrow corresponds to the direction of transcription. Hairpin loops with a stem length of more than 7 bp and a loop of 3 to 20 bases are marked by little black arrows. ORFs with adjacent numbers are members of corresponding gene families defined by significant homology to one another.

TABLE 1. Supposed THV coding ORFs[a]

THV strain 2 ORF	Nucleotide position	Length (aa)	Potential site and signature (position within protein sequence)	Most-homologous protein(s) (no. of aa); possible function	% I/S
T1 (R)	258–1811	518	Mitochondrial energy transfer protein signature (154–162); cereal trypsin–α-amylase inhibitor family signature (207–230)	HCMVA US23 (592), MCMVS M143 (557); transactivator	23.0/34.8, 21.4/42.3
T2 (R)	2034–4220	729	4 bipartite nuclear targeting sequences (459–475; 460–476; 461–477; 462–478); leucine zipper pattern (244–265)	HCMVA US23 (592), MCMVS M143 (557); transactivator	25.0/29.4, 21.5/33.2
T3 (R)	4486–5973	496	Cell attachment sequence (372–374)	HCMVA US24 (500), MCMVS M141 (508); transactivator	32.6/29.5, 25.6/29.5
T4 (R)	6111–7685	525		HCMVA US23 (592), MCMVS M140 (484); transactivator	32.6/32.1, 29.8/37.4
T5 (R)	7917–9890	658	Mitochondrial energy transfer protein signature (252–260)	HCMVA US22 (593), MCMVS M139 (644); transactivator	29.4/32.7, 23.6/32.9
t6 (L)	9998–10552	185			
t7 (R)	10231–12087	619			
t8 (R)	12154–13281	376			
t9 (R)	13387–14208	274	Cell attachment sequence (227–229)		
t10 (L)	13488–14963	492	Bipartite nuclear targeting sequence (426–442)		
t11 (R)	14252–14950	233	Microbody C-terminal targeting signal (231–233)		
t12 (L)	14872–15195	108	Bipartite nuclear targeting sequence (74–90)		
t13 (R)	15350–15532	61	Bipartite nuclear targeting sequence (31–47)		
t13.1 (R)	15632–15952	107			
t14 (L)	16283–16930	216			
t15 (R)	17685–17963	93			
t16 (L)	18440–19675	412	Aminoacyl-transfer RNA synthetase class II signature (170–179)		
t16.1 (R)	19060–19773	238			
t17 (L)	19807–20799	331			
t18 (R)	20975–22045	357			
t19 (L)	22166–22444	93			
t20 (L)	22656–23621	322			
t21 (L)	24376–25323	316			
t22 (L)	25743–26816	358			
t22.1 (L)	26749–27849	367			
t22.2 (L)	28262–29341	360			
t22.3 (L)	29622–30719	366			
t22.4 (L)	30811–30969	53			
t22.5 (L)	31313–31855	181			
t22.6 (L)	32271–32684	138			
t22.7 (L)	33015–34052	346			
t22.8 (L)	34304–35407	368			
t22.9 (R)	35289–35702	138	Cell attachment sequence (51–53)		
t22.10 (L)	35795–36220	142			
t22.11 (R)	36085–36486	134			
t22.12 (L)	36591–36791	67			
t22.13 (R)	36656–36835	60			
T23 (L)	37706–38629	308		HCMVA UL23 (342), MCMVS M23 (391); transactivator	25.9/29.5, 24.6/31.7
T25 (R)	39053–41116	688		HCMVA UL25 (656), MCMVS M25 (932); tegument protein	20.5/39.3, 17.8/35.6
T26 (L)	41208–41933	242		HCMVA UL26 (188), MCMVS M26 (192); virion protein	27.3/33.1, 25.3/31.0
T27 (L)	42066–43933	636	Bipartite nuclear targeting sequence (266–282); multicopper oxidase signature 1 (164–184)	HCMVA UL27 (608), MCMVS M27 (682)	28.2/34.4, 25.4/38.7
T27.1 (R)	43846–44466	207	Bipartite nuclear targeting sequence (159–175); Cell attachment sequence (36–38)		
T28 (L)	44135–45157	341		HCMVA UL29 (360), MCMVS M28 (430); transactivator	17.9/35.1, 26.8/27.7
t28.1 (L)	45174–46217	348	Gram-positive coccus surface protein anchoring hexapeptide (162–167)		
t28.2 (L)	46395–47030	212	Microbody C-terminal targeting signal (210–212)		
T29 (L)	46951–48744	598	3 bipartite nuclear targeting sequences (384–400; 385–401; 516–532)	HCMVA UL29 (360); transactivator	21.7/22.1
T29.1 (R)	47212–47598	129	Cell attachment sequence	MCMVS M106 (147)	22.4/42.9
t29.2 (R)	48411–49493	361	Bipartite nuclear targeting sequence (236–252)		
t29.3 (L)	49402–50247	282			
T30 (L)	50424–52073	550		HSV6 VU4	20.5/38.2
T31 (R)	51785–53419	545		HCMVA UL31 (694), MCMVS M31 (516)	21.0/30.5, 14.6/40.9
T32 (L)	53284–55437	718	Leucine zipper pattern (193–214); pfkB family of carbohydrate kinase signature 1 (326–349)	HCMVA UL32 (1048), MCMVS M32 (718); major tegument protein	18.3/32.1, 21.0/39.2
T32.1 (R)	54701–55078	126		MCMVS M19 (147)	20.4/28.3

Continued on facing page

TABLE 1—*Continued*

THV strain 2 ORF	Nucleotide position	Length (aa)	Potential site and signature (position within protein sequence)	Most-homologous protein(s) (no. of aa); possible function	% I/S
T33 Exon 1 (R) Exon 2 (R)	55618–55650 55754–56815	365	GCR signature (119–135)	HCMVA UL33 (390), MCMVS M33 (377); GCR homologue	39.2/30.3, 51.6/33.6
T34 (R)	57213–58445	411		HCMVA UL34 (504), MCMVS M34 (854)	21.5/36.4, 15.8/22.9
T35 (R)	58631–60331	567		HCMVA UL35 (640), MCMVS M35 (519)	26.5/35.2, 30.9/38.3
T36 (L)	60733–62329	?	?	HCMVA UL36 (476), MCMVS M36 (507); immediate-early regulatory protein	?
T37 (L)	62518–63612	365		HCMVA UL37 (932), MCMVS M37 (345); immediate early regulatory protein	17.2/31.3, 22.8/39.9
T38 (L)	63903–65411	503		HCMVA UL38 (331), MCMVS M38 (497); immediate-early regulatory protein	22.8/25.5, 22.6/40.7
t39 (R)	65237–65635	133	Microbody C-terminal targeting signal (131–133); bipartite nuclear targeting sequence (99–115)		
t40 (L)	65341–66024	228			
t41 (R)	66103–66984	294	Bipartite nuclear targeting sequence (15–31); cell attachment sequence (22–24)		
t42 (L)	66371–66931	187			
t43 (L)	67053–67781	243			
t43.1 (R)	67950–68267	106			
t43.2 (L)	68122–68595	158			
t43.3 (L)	68598–68804	69			
t43.4 (L)	69102–70094	331			
T44 (L)	70088–71233	382		HCMVA UL44 (433), MCMVS M44 (411); polymerase accessory protein	53.7/29.5, 54.1/30.4
T45 (L)	71625–74729	1,035	Bipartite nuclear targeting sequence (8–24)	HCMVA UL45 (906), MCMVS M45 (876); ribonucleotide reductase	22.8/36.0, 26.0/37.4
T46 (L)	74745–75692	316		HCMVA UL46 (290), MCMVS M46 (294); capsid assembly and maturation protein	38.1/37.1, 39.5/32.6
T47 (R)	75692–78856	1,055	Leucine zipper pattern (672–693)	HCMVA UL47 (982); MCMVS M47 (1040); capsid assembly protein	29.6/40.3, 24.9/40.2
T48 (R)	78881–85765	2,295	Bipartite nuclear targeting sequence (2184–2200); cell attachment sequence (936–938); 3 leucine zipper patterns (431–452, 2260–2281, 2267–2288); sugar transport protein signature 1 (1610–1623)	HCMVA UL48 (2241), MCMVS M48 (2149); large tegument protein	30.8/16.4, 29.4/16.6
t48.1 (R)	86101–86412	104	Prenyl group binding site (CAAX box 101–104); Bowman-Birk serine protease inhibitor family signature (94–101)		
T49 (L)	86568–88331	588	Bipartite nuclear targeting sequence (250–266); cell attachment sequence (122–124)	HCMVA UL49 (570), MCMVS M49 (536); viral protein	40.3/29.6, 40.6/30.0
T50 (L)	88303–89322	340		HCMVA UL50 (397), MCMVS M50 (316); viral protein	40.5/31.2, 41.1/33.1
T51 (L)	89539–89895	119		HCMVA UL51 (157), MCMVS M51 (233); DNA cleavage and packaging protein	38.9/27.4, 29.6/15.9
T52 (R)	89985–91709	575		HCMVA UL52 (668), MCMVS M52 (517); major envelope glycoprotein	37.9/33.5, 45.1/32.2
T53 (R)	91705–92643	313		HCMVA UL53 (376), MCMVS M53 (333); Viral protein	37.8/25.4, 43.7/35.4
T54 (L)	92738–96250	1,171	DNA polymerase family B signature (873–881)	RHCMV UL54, HCMVA UL54 (1,243), MCMVS M54 (1,097); DNA polymerase	50.7/26.5, 43.5/25.5, 48.5/26.6
T55 (L)	96253–99084	944		HCMVA UL55 (907), MCMVS M55 (937); glycoprotein B	46.2/33.7, 41.6/35.8
T56 (L)	98906–101251	782		HCMVA UL56 (851), MCMVS M56 (798); probable processing and transport protein	51.1/30.1, 54.1/32.0
T57 (L)	101550–105128	1,193	Leucine zipper pattern (497–518); aldehyde dehydrogenase cysteine active site (1087–1098); crystalline beta and gamma Greek key motif signature (809–824)	HCMVA UL57 (1,236), MCMVS M57 (1,191); major DNA-binding protein	54.1/30.1, 52.7/33.7
t58 (R)	104869–105630	254	2 cell attachment sequences (212–214, 217–219)		
t59 (L)	105320–105982	221			
t60 (R)	108366–108890	175	4 bipartite nuclear targeting sequences (120–136, 133–149, 134–150, 135–151); 7 leucine zipper pattern (44–65, 51–72, 58–79, 65–86, 72–93, 79–100, 86–107)		

Continued on following page

TABLE 1—*Continued*

THV strain 2 ORF	Nucleotide position	Length (aa)	Potential site and signature (position within protein sequence)	Most-homologous protein(s) (no. of aa); possible function	% I/S
t61 (L)	109040–110077	346			
t62 (R)	110625–110900	92	Bipartite nuclear targeting sequence (10–26)		
t63 (L)	110625–110861	79			
T69 (L)	111240–113507	756		HCMVA UL69 (744), MCMVS M69 (841); transcriptional regulator	23.6/44.8; 22.8/36.4
t69.1 (L)	113441–113857	139			
T70 (L)	113758–116667	970	Leucine zipper pattern (642–663)	HCMVA UL70 (1062), MCMVS M70 (964); helicase-primase complex protein	38.3/27.9, 39.5/31.9
T71 (R)	116762–118072	437		HCMVA UL71 (411), MCMVS M71 (299); viral protein	27.0/32.0, 26.0/25.6
T72 (L)	117603–118838	412	Bipartite nuclear targeting sequence (389–405)	HCMVA UL72 (388), MCMVS M72 (382); dUTP-pyrophosphatase	24.4/37.4, 20.0/35.5
T73 (R)	118838–119161	108		HCMVA UL73 (138), MCMVS M73 (94); membrane protein	27.5/34.1, 34.5/37.3
T74 (L)	119153–120544	464		HCMVA UL74 (466), MCMVS M74 (438); glycoprotein H-L complex component	19.8/41.2, 16.9/39.6
Y75 (L)	120745–122934	730	Prenyl group binding site (CAAX box; 727–730)	HCMVA UL75 (743), MCMVS M75 (725); glycoprotein H	30.3/39.8, 30.2/38.8
T76 (R)	123058–123924	289	3 bipartite nuclear targeting sequences (267–283, 268–284, 269–285)	HCMVA UL76 (325), MCMVS M76 (254); viral protein	36.4/28.1, 37.7/25.7
T77 (R)	123662–125401	580	Bipartite nuclear targeting sequence (29–45)	HCMVA UL77 (642), MCMVS M77 (628); DNA cleavage and packaging protein	47.8/28.4, 45.1/31.1
T78 (R)	125593–126984	464		HCMVA UL78 (431), MCMVS M78 (471); GCR homologue	21.4/37.0, 16.4/42.4
T78.1 (L)	126986–127720	245	Bipartite nuclear targeting sequence (69–85)	MCMVS M59 (340)	20.8/32.7
T79 (L)	127033–127845	271		HCMVA UL79 (295), MCMVS M79 (258); viral protein	45.9/28.7, 49.4/28.4
T80 (R)	127913–130114	734		HCMVA UL80 (708), MCMVS M80 (697); scaffolding protein: protease	31.3/38.3, 33.1/38.4
T82 (L)	130452–132365	638	Microbody C-terminal targeting sequence (636–638); cell attachment sequence (86–88)	HCMVA UL82 (559), MCMVS M82 (598); tegument phophoprotein	24.1/35.8, 18.5/40.3
T84 (L)	132502–133935	478	Leucine zipper pattern (116–137, 123–144)	HCMVA UL84 (586), MCMVS M84 (587); DNA replication regulatory protein	22.3/34.0, 17.3/36.5
T85 (L)	134070–134993	308		HCMVA UL85 (306), MCMVS M85 (311); capsid protein	58.3/27.2, 53.3/34.3
T86 (L)	135032–139189	1,386	Cell attachment sequence (1292–1294)	HCMVA UL86 (1,370), MCMVS M86 (1,353); major capsid protein	57.9/15.2, 57.4/14.5
T87 (R)	139257–141196	980	Cell attachment sequence (368–370)	HCMVA UL87 (941), MCMVS M87 (926); viral protein	41.6/30.2, 45.5/28.7
T88 (R)	142066–143223	386	Cell attachment sequence (158–160); leucine zipper pattern (333–354)	HCMVA UL88 (429), HCMVA M88 (426); virion protein	29.0/34.9, 27.5/34.6
T89E2 (L)	143242–144366	375		HCMVA UL89 (674), MCMVS M89 (671); DNA cleavage and packaging protein	66.4/24.9, 68.9/23.8
t90 (L)	144657–145304	216	GCR signature (51–67)		
T91 (R)	144758–145099	114		HCMVA UL91 (111), MCMVS M91 (134); viral protein	33.3/47.4, 22.8/22.8
T92 (R)	145145–145831	229	Bipartite nuclear targeting sequence (149–165)	HCMVA UL92 (201), MCMVS M92 (230); viral protein	48.5/28.8, 58.7/26.5
T93 (R)	145800–147422	541	Cell attachment sequence (77–79)	HCMVA UL93 (594), MCMVS M93 (515); viral protein	23.3/32.8, 28.1/37.8
T94 (R)	147368–148440	361		HCMVA UL94 (345), MCMVS M94 (345); capsid-tegument protein	33.5/35.1, 34.7/38.8
T89E1 (L)	148569–149462	298		HCMVA UL89 (674), MCMVS M89 (671); DNA cleavage and packaging protein	66.4/24.9, 68.9/23.8
T96 (R)	150689–151063	125		HCMVA UL96 (115), MCMVS M96 (129); viral protein	22.8/44.9, 26.8/41.3
T97 (R)	151128–153176	683	Tyrosine protein kinase-specific active-site signature (407–419)	HCMVA UL97 (707), MCMVS M97 (643); phosphotransferase	32.2/36.3, 29.6/34.3
T98 (R)	153341–155056	572		HCMVA UL98 (584), MCMVS M98 (561); alkaline exonuclease	43.1/31.0, 33.2/34.1
T99 (R)	154996–155439	148		HCMVA UL99 (112), MCMVS M99 (190); tegument phosphoprotein	18.6/43.8, 24.3/33.1
T100 (L)	156038–157090	351		HCMVA UL100 (372), MCMVS M100 (371); glycoprotein M	49.5/30.4, 50.7/30.9

Continued on facing page

TABLE 1—*Continued*

THV strain 2 ORF	Nucleotide position	Length (aa)	Potential site and signature (position within protein sequence)	Most-homologous protein(s) (no. of aa); possible function	% I/S
t101 (R)	157036–157410	125	Bipartite nuclear targeting sequence (50–66)		
t101.1 (L)	157060–157386	109	Bipartite nuclear targeting sequence (13–29)		
T102 (R)	157386–159761	792	Leucine zipper pattern (337–358)	HCMVA UL102 (798), MCMVS M102 (812); glycoprotein M	26.3/34.9, 22.6/35.3
T103 (L)	160023–160808	262		HCMVA UL103 (249), HCMVA M103 (317); viral protein	34.3/38.5, 35.5/35.5
T104 (L)	160762–162927	722	Bipartite nuclear targeting sequence (687–703)	HCMVA UL104 (697), MCMVS M104 (704); DNA cleavage and packaging protein	45.5/31.5, 40.4/35.4
T105 (R)	162774–165668	965	ATP-GTP-binding site motif A (P loop) (176–183)	HCMVA UL105 (956), MCMVS M105 (948); (helicase)	52.1/25.8, 47.5/25.4
t106 (L)	165625–166041	139			
t107 (R)	167264–167497	78			
t108 (L)	168103–168495	131			
t109 (L)	168149–168592	148			
t110 (R)	168666–168983	106			
t111 (R)	168913–169239	109			
t111.1 (R)	169430–169777	116			
t111.2 (R)	170547–170786	80	Cell attachment sequence (75–77)		
t111.3 (L)	172034–172606	191			
t112 (R)	172345–173073	243		HCMVA UL112 (268), MCMVS M112EI (264); (early phosphoprotein)	33.5/33.5, 32.2/35.1
T114 (L)	174794–175612	273		HCMVA UL114 (250), MCMVS M114 (262); (uracil-DNA glycosylase)	54.4/21.5, 50.9/23.7
T115 (L)	175578–176591	338		HCMVA UL115 (278), MCMVS M115 (274); (glycoprotein L)	32.8/28.0, 33.9/32.4
t116 (L)	176599–177567	323			
T117 (L)	177626–178804	393		HCMVA UL117 (424)	20.9/45.3
t118 (R)	178851–179195	115	Leucine zipper pattern (29–50)		
t119 (L)	179674–180630	319	ATP-dependent DNA ligase AMP-binding site (213–221)		
t120 (R)	180782–181438	219	2 bipartite nuclear targeting sequences (136–152, 137–153)		
t121 (L)	181046–181849	268			
t121.1 (R)	181092–182258	389	2 bipartite nuclear targeting sequences (16–32, 373–389)		
t121.2 (R)	182632–183879	416	4 bipartite nuclear targeting sequences (313–329, 314–330, 320–336, 368–384); 2 cell attachment sequences (204–206, 214–216)		
t121.3 (L)	182778–184280	501	2 bipartite nuclear targeting sequences (99–115, 100–116); 2 cell attachment sequences (9–11, 388–390)		
t121.4 (L)	183833–184690	286	Prenyl group binding site (CAAX box; 283–286)		
t121.5 (L)	184602–185216	205			
T122 (L)	186729–187973	415		HCMVA UL122 (411), MCMVS M122E5 (511); immediate-early regulatory protein	33.2/44.0, 25.4/33.3
t123 (L)	188849–189949	367			
t124 (R)	190929–191357	143			
t125 (L)	192003–192191	63			
t126 (L)	193041–193211	57			
t127 (L)	193423–193854	144	2 bipartite nuclear targeting sequences (33–49,		
t128 (L)	194052–194315	88			
t129 (R)	194386–194718	111			
t130 (R)	194655–195239	195	Bipartite nuclear targeting sequence (82–98)		

[a] aa, amino acids. %I/S percentage of amino acids identical (I) and similar (S) to the corresponding homologous proteins. RHCMV, rhesus cytomegalovirus. The homology values of T89 exons 1 and 2 are given for the complete corresponding proteins of HCMVA and MCMVS. Genetic analysis revealed that T36 is spliced, but it was not possible to prove this by RT-PCR experiments. For this reason, the length and homology values of T36 are replaced by question marks. The proteins that are most homologous to the individual potential THV proteins are underlined. The individual ORFs are numbered in the order in which they appear in the THV genome with the designation L for theoretical transcription to the left or R for theoretical transcription to the right. The sites and signatures found in the amino acid sequences of the potential THV proteins were determined by the program PROSITE of the PC/GENE software package and are putative. Shown are all ORFs that are longer than 300 bp and all ORFs that are longer than 150 bp if they are positioned between ORFs of more than 300 bp and that are supposed to code for viral proteins according to accepted rules for the choice of transcribed herpesvirus ORFs (see Materials and Methods). The THV ORFs are designated T (homologous to HCMVA and/or MCMVS proteins) or t (no homologous protein detected). The T THV ORFs are numerated according to the homologous proteins of HCMVA and MCMVS; the t THV ORFs are numbered to fill the gaps between the T THV ORFs.

coding sequences led to the identification of six gene families. The members of each of these six gene families are distinguished by significant identity and similarity to each other, raising the possibility that these gene groups in each case originated from one ancestral gene by duplication events. The gene families are numbered 1 to 6 and drawn in the ORF map in Fig. 2. A striking feature of the members of such gene groups is the fact that they are located in close proximity in the

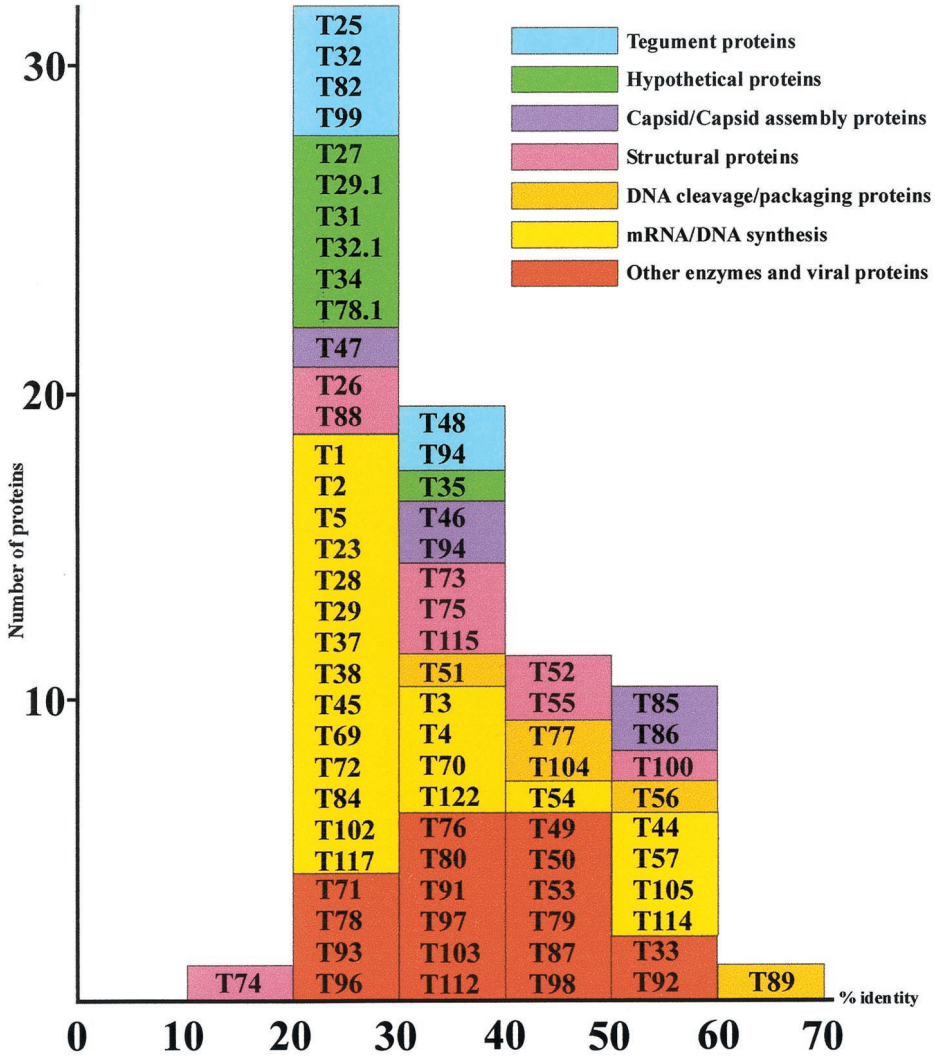


FIG. 3. Graphic representation of the homology value distribution of seven conserved THV protein groups that are classified according to possible functions. Individual groups are marked by distinct colors. The homology values that underlie this graphic are the highest detected homologies to known proteins regardless of the virus species the proteins belong to.

genome, often in a tandem-like arrangement, underlining the hypothesis that they were produced by gene duplication. A concentration of gene family members could be seen in the left part of the THV genome, especially in the nonconserved region between t6 and t22.13. A similar distribution of gene families is present in the genomes of HCMVA (13), MCMVS (71), rat cytomegalovirus (87), or HHV-6 (31), where duplicated genes are also concentrated in nonconserved regions at the left and right ends of the genomes. Significant homology between the members was the main criterion for the formation of the six THV gene families. Another possible way to relate genes to families is by the similarity of the functions of the encoded proteins. In HCMVA and MCMVS, homologues of G protein-coupled receptors (GCRs) form a family. For that reason, T33 and T78, both homologous to G protein-coupled receptors, could also be seen as members of a gene family. This GCR and members of the US22 gene family are conserved among THV, HCMVA, and MCMVS. Both members of the HCMVA UL25-UL35 gene family are conserved in the THV genome. However, homology between the two potential proteins is extremely low.

Gene arrangements of the THV genome. Homologous THV, HCMVA, and MCMVS proteins are summerized in Table 1. All together, 72 proteins are homologous between THV and HCMVA and 73 are homologous between THV and MCMVS. Both cytomegalvirus species are members of the subfamily *Betaherpesvirinae*. Table 2 shows these homology relationships extended by HHV-6 (*Betaherpesvirinae*) (31), EBV (*Gammaherpesvirinae*) (4), and HSV-1 (*Alphaherpesvirinae*) (58, 59, 60). HHV-6 shares 66, EBV shares 42, and HSV-1 shares 38 homologous proteins with THV. The most homologous proteins are found between THV and the mammalian cytomegaloviruses. The extent of homology between THV and alpha- or gammaherpesviruses is clearly less. Furthermore, Table 2 shows a group of almost 40 homologous proteins that could be found in every herpesvirus species of mammals and birds. These genes are termed core genes and are supposed to be part of the genome of the common ancestor of these herpesviruses. The

TABLE 2. HCMVA, MCMVS, HHV-6, EBV, and HSV-1 proteins homologous to potential THV proteins[a]

THV protein	HCMVA homologue	MCMVS homologue	HHV-6 homologue	EBV homologue	HSV-1 homologue
T1 (US22 family)	US23	M143	U16		
T2 (US22 family)	US23	M143	U16		
T3 (US22 family)	US24	M141			
T4 (US22 family)	US23	M140	U16		
T5 (US22 family)	US22	M139	DR7		
T23	UL23	M23	U2		
	UL24	M24	U3		
T25	UL25	M25	U14		
T26	UL26	M26			
T27	UL27	M27	U4/U5		
T28 (US22 family)	UL29	M28	U8		
T29 (US22 family)	UL29		U8		
T29.1		M106			
T30	UL27	M27	U4/U5		
T31	UL31	M31	U10		
T32	UL32	M32	U11		
T32.1		M19	U12		
T33	UL33	M33		EB11	
T34	UL34	M34			
T35	UL35	M35			
T36 (US22 family)	UL36	M36			
T37	UL37	M37	U18		
T38	UL38	M38	U19		
	UL43	M43			
T44	UL44	M44	U27	BMRF1	UL42
T45	UL45	M45	U28	BORF2	UL39
T46	UL46	M46	U29	BORF1	UL38
T47	UL47	M47	U30	BOLF1	UL37
T48	UL48	M48	U31	BPLF1	UL36
T49	UL49	M49	U33	BFRF2	UL35
T50	UL50	M50	U34	BFRF1	UL34
T51	UL51	M51	U35		UL33
T52	UL52	M52	U36	BFLF1	UL32
T53	UL53	M53	U37	BFLF2	UL31
T54	UL54	M54	U38	BALF5	UL30
T55	UL55	M55	U39	BALF4	UL27
T56	UL56	M56	U40	BALF3	UL28
T57	UL57	M57	U41	BALF2	UL29
T69	UL69	M69	U42	BMLF1	UL54
T70	UL70	M70	U43	BSLF1	UL52
T71	UL71	M71	U44	BSRF1	UL51
T72	UL72	M72	U45	BLLF2	UL50
T73	UL73	M73	U46	BLRF1	UL49A
T74	UL74	M74	U47		
T75	UL75	M75	U48	BXLF2	UL22
T76	UL76	M76	U49	BXRF1	UL24
T77	UL77	M77	U50	BVRF1	UL25
T78	UL78	M78	U51		
T78.1		M59			
T79	UL79	M79	U52	BVRF1.5a/b	
T80	UL80	M80	U53	BVRF2	UL26
T82	UL82	M82	U54		
	UL83	M83			
T84	UL84	M84	U55		
T85	UL85	M85	U56	BDLF1	UL18
T86	UL86	M86	U57	BcLF1	UL19
T87	UL87	M87	U58	BcRF1	
T88	UL88	M88	U59		
T89	UL89	M89	U60/U66	BDRF1/BGRF1	UL15Ex2+1
T91	UL91	M91	U62		
T92	UL92	M92	U63	BDLF4	
T93	UL93	M93	U64	BGLF1	UL17
T94	UL94	M94	U65	BGLF2	UL16
T95	UL95	M95	U67	BGLF3	UL14
T96	UL96	M96	U68	BGLF3.5	
T97	UL97	M97	U69	BGLF4	UL13
T98	UL98	M98	U70	BGLF5	UL12
T99	UL99	M99			
T100	UL100	M100	U72	BBRF3	UL10
T102	UL102	M102	U74	BBLF3	UL8
T103	UL103	M103	U75	BBRF2	UL7
T104	UL104	M104	U76	BBRF1	UL6
T105	UL105	M105	U77	BBLF4	UL5
T112	UL112	M112E1	U79		
T114	UL114	M114	U81	BKRF3	UL2
T115	UL115	M115	U82	BKRF2	UL1
	UL116	M116			
T117	UL117		U84		
	UL118	M118			
	UL121	M121			
T122	UL122	M122E5	U86		

[a] The data shown are from reference 31 and are complemented by THV and MCMVS protein data.

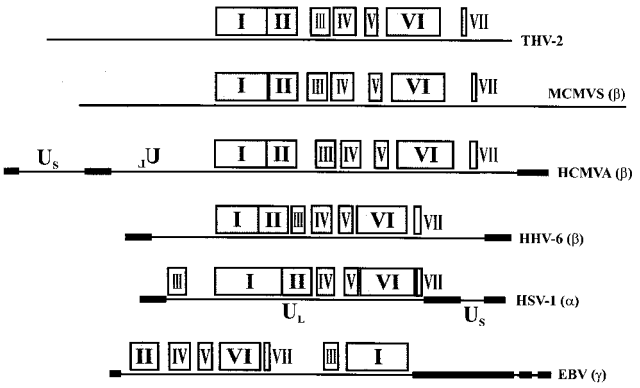


FIG. 4. Arrangement of the seven conserved core gene blocks in herpesvirus genomes. The pattern of these seven clusters is characteristic for the three herpesvirus subfamilies. The assignment of the individual virus species to the subfamily *Alpha*-, *Beta*-, or *Gammaherpesvirinae* is given in parentheses after the particular virus name. U_s (unique short region) and U_L (unique long region) are the descriptions of the two parts of the herpesvirus genomes that are able to form isomers. The inverted spelling of U_L in the genome of HCMVA characterizes a distinct isomer that differs from the prototype L-S isomer in the orientation of the U_L region. The black boxes mark the positions of large repetitive elements. The diagram corresponds to the 1996 publication of Gompels et al. (31) complemented by data on the THV and MCMV genomes.

core genes form seven conserved gene clusters whose arrangements are characteristic of the corresponding herpesvirus subfamily. A comparison of the gene block arrangements among THV, HCMVA, MCMVS, HHV-6, HSV-1, and EBV is shown in Fig. 4. In the genomes of THV, MCMVS, and HCMVA, the arrangement of the seven clusters is almost the same with regard to both the order of appearance and the spaces between the individual blocks. The same order of gene clusters could also be found in HHV-6, but they are more concentrated in the center of the genome. HCMVA, MCMVS, and HHV-6 are members of the subfamily *Betaherpesvirinae*. The gene cluster arrangement in the genome of HSV-1, and alphaherpesvirus, differs from those of the betaherpesviruses in the position of gene block III, which is localized at the left end of the U_L region of the HSV-1 genome. The gammaherpesvirus EBV shows a further different arrangement that is characterized by the localization of gene blocks III and I at the right end of the long unique genome region of EBV. The colinear arrangement of conserved genes within the genomes of THV, HCMVA, and MCMVS is plotted in detail in Fig. 5. HCMVA is able to form four genomic isomers. The arrangement of genes in the prototype HCMVA isomer is colinear with that in the MCMVS genome, which is not able to form isomers. The arrangement and orientation of the homologous THV genes correspond to those of a different HCMVA genome isomer that is characterized by an inverted U_L region compared to the prototype HCMVA isomer. The difference in total genome length between THV (195,857 bp) and HCMVA (229,354 bp) is caused mainly by the presence of the repetitive elements in the HCMVA genome. Without these elements, the lengths of the THV and HCMVA genomes are very similar.

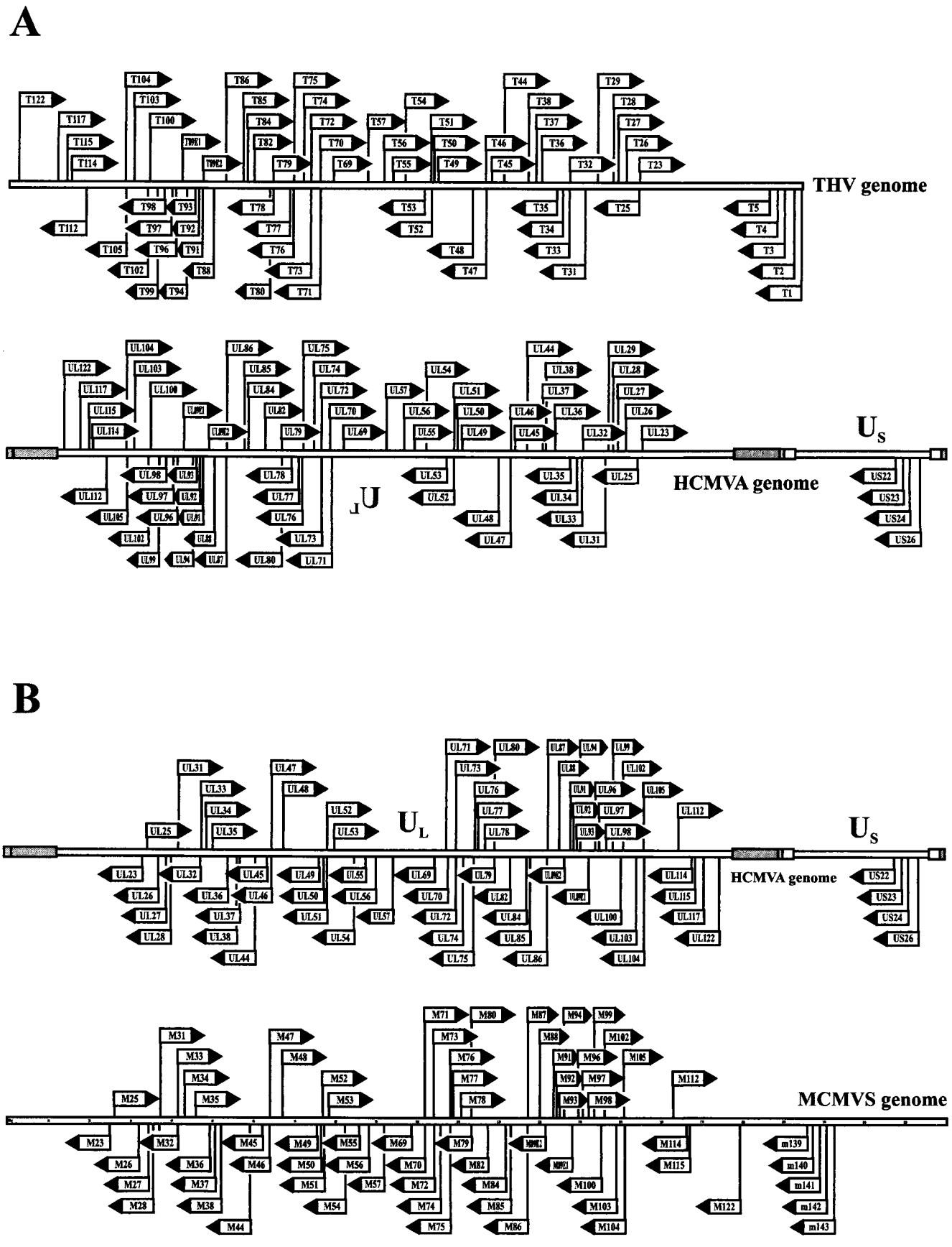
Phylogenetic classification of THV. The phylogenetic trees derived from the comparison of the DNA polymerase, DNA polymerase accessory protein, glycoprotein B, probable transport and processing protein, major DNA-binding protein, major capsid protein, viral terminase, and uracil DNA-glycosylase amino acid sequences of different herpesviruses are shown in Fig. 6A to H. The selected proteins are those with the highest levels of homology between different members of the family *Herpesviridae*. The phylogenetic trees show a distinct subdivision into three main branches corresponding to the herpesvirus subfamilies *Alpha*-, *Beta*-, and *Gammaherpesvirinae*, which are groups of herpesvirus species with similar biological properties and phylogenetic relatedness. In all eight trees, THV is a member of the subfamily *Betaherpesvirinae*. In addition, the evolutionary lineage of this subfamily is divided into two branches that correspond to the genus *Roseolovirus* with HHV-6 and HHV-7, the so-called Beta2 herpesviruses, and the mammalian cytomegaloviruses, the so-called Beta1 herpesviruses. Within the subfamily *Betaherpesvirinae*, THV is most closely related to the mammalian cytomegaloviruses, with similar evolutionary distances to the phylogenetic lineages of primate and rodent cytomegaloviruses.

DISCUSSION

The complete nucleotide sequence (195,857 bp) and the coding capacity of the THV genome were determined. The position of THV as the prototype species of herpesvirus genome group F (49, 72) was confirmed. The G+C content (66.5%) of the THV genome varies over the course of the genome. The highest values were found at the termini of the viral genome and within the DNA sequences of the conserved genes. The mechanism for G+C accumulation is not known. It occurs in the genomes of different herpesvirus species (e.g., EBV [4] or HSV-1 [58, 59, 60]) regardless of phylogenetic relatedness and is supposed to be an adaptation to an unknown evolutionary pressure.

Seventy-six out of 158 potential gene products of THV were identified as significantly homologous to known herpesvirus proteins of primate and rodent cytomegaloviruses, mainly those of HCMVA and MCMVS. The thorough examination of many of these homologous proteins allows the assignment of functions to the potential THV proteins. T1 to T5, T28, T29, and T36 are homologues of the US22 gene family of HCMVA. The members of the US22 gene family are known to regulate

FIG. 5. Comparative analysis of the arrangement of homologous ORFs among THV, HCMVA, and MCMVS. Each arrow shows the orientation of transcription. The length of each arrow is standardized and does not correspond to the actual length of the ORF. The vertical lines designate the start points of the individual ORFs. The shaded boxes of the HCMVA genome represent repetitive DNA elements that are, in part, responsible for genomic isomerization. These repetitive DNA elements divide the HCMVA genome into U_L (unique long) and U_s (unique short) regions. (A) Comparison of the arrangement of homologous ORFs between THV and HCMVA. The THV ORFs show colinearity with those of a distinct HCMVA genome isomer that is characterized by an inverted U_L region compared to the prototype L-S isomer. (B) Comparison of the arrangement of homologous ORFs between HCMVA and MCMVS. The MCMVS ORFs show colinearity with those of the prototype HCMVA genome isomer.


A
THV genome
HCMVA genome
U_S
U_L
B
U_L
U_S
HCMVA genome
MCMVS genome

gene expression (13). m139 to m143 are the corresponding MCMVS homologues to THV T1 to T5. They are supposed to play an essential role in pathogenicity in the natural host and seem to be important for genome replication (35). HCMVA UL122 and UL123 are the main components of the so-called major immediate-early region (32, 45, 83, 84). These genes correspond to M122, M123, and T122 of MCMVS (12, 65) and THV, respectively. The mRNAs of UL122, UL123, M122, and M123 consist of several exons composed by alternative splicing events. It is very probable that the corresponding THV genes have similar exon structures. However, the transcription of these viral genes will be the subject of future studies. HCMVA immediate-early genes play an essential role in the regulation of the expression of early and late viral genes and are indispensable for the correct course of the lytic replication cycle. THV T36 to T38 and T115 are homologous to HCMVA immediate-early proteins UL36 to UL38 and UL115 (14, 15). UL37 is an integral membrane protein (1) that is supposed to be located in mitochondria and to inhibit Fas-mediated apoptosis of the host cell (30).

Herpesvirus tegument proteins have structural functions in viral morphogenesis and are involved in the regulation of gene expression immediately after penetration of host cells. In addition, some of them are responsible for the activation of the cellular immune response (33). THV gene products T25, T32, T47, T48, T69, T82, and T99 are homologous to tegument proteins of HCMVA and MCMVS (6, 16, 17, 18, 37, 38, 64, 91, 94). HCMVA UL32, UL82, and UL83 are the main components of the viral tegument. UL32 was designated the major tegument phosphoprotein and makes up 15% of the total protein mass of a virion and plays an important role in viral morphogenesis (34, 66). HCMVA UL83 is essential for the viral life cycle in the natural host but not in cell culture (77). A homologue to UL83 is present in MCMVS (71) and rat cytomagalovirus (87) but absent in HHV-6 (31) and THV. HCMVA UL69, which is homologous to THV T69, has been supposed to be a transactivator (90) and plays a role in the G_1 phase of the host cell cycle (36, 55).

THV T46 (minor capsid protein), T85, T86 (major capsid protein), and T94 are homologous to the corresponding capsid proteins of HCMVA and MCMVS (13, 29, 71, 89). The transcription unit of the smallest capsid protein of HCMVA is 225 bp in length and located between UL48 and UL49 (28). A homologous protein in THV could not be detected. However, a small ORF (t48.1) of 312 bp located between ORFs T48 and T49 has, moreover, a potential Bowman-Birk serine protease inhibitor family signature. THV T80 is the homologue of the viral protease, one of the most important herpesvirus enzymes, which is a scaffolding protein with essential functions in the morphogenesis of the viral capsid (7, 11, 54, 75, 92).

Eleven gene loci were identified within the HCMVA genome, which are essential for DNA replication. UL54 (DNA polymerase), UL44 (DNA polymerase accessory protein), UL57 (major DNA-binding protein), UL70 (primase), UL102 (primase-helicase complex-associated protein), and UL105 (helicase) (46, 70, 80, 81) are replication fork proteins and are also present in THV. T54, the DNA polymerase of THV, is highly conserved and possesses characteristic sites and signatures of the B family of DNA polymerases (82). T57 shows a similar strong conservation and possesses sequence motifs essential for DNA binding (5). Homologues of THV T36-38, T84, T112, and T122 are also essential for DNA replication. They are supposed to activate the expression of the replication fork genes (27, 76, 79).

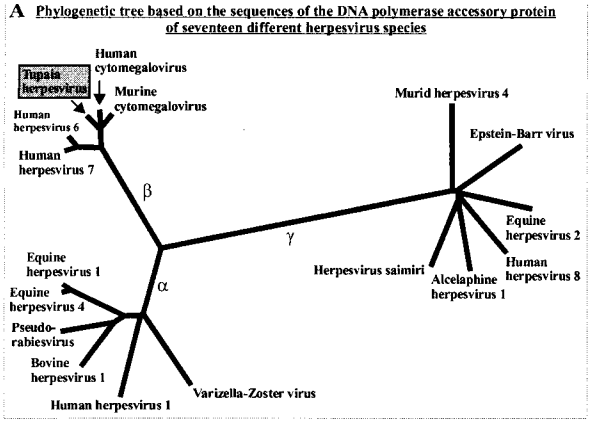
In the course of herpesvirus DNA replication, which proceeds by the rolling-circle mechanism, concatemeric genomes are formed (50). HCMVA UL51, UL52, UL56, UL77, UL89, and UL104 were identified as the vital components of the cleavage and packaging processes that are essential for the formation of unit length viral genomes and correct packaging (51). THV possesses individual homologues to these six proteins.

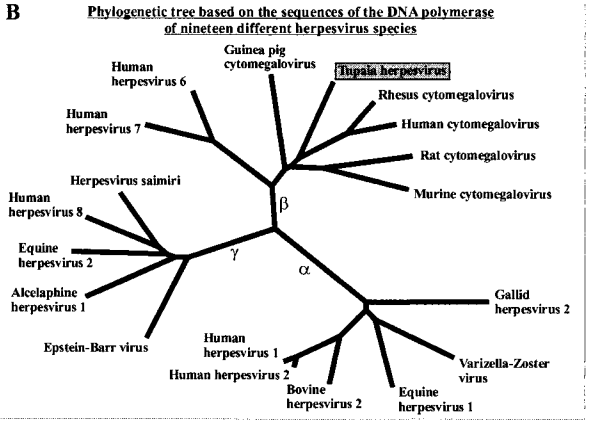
THV homologues of nucleic acid metabolism are T45, T72 (dUTPase), and T114 (uracil DNA-glycosylase). T45 is the large chain of the ribonucleotide reductase. All betaherpesviruses, including THV, have no homologous ORF for the small chain of the ribonucleotide reductase that actually possesses the active site of the enzyme. It is not known how the betaherpesvirus ribonucleotide reductases retain their function. Like other G+C-rich herpesviruses like EBV (4) or HSV-1 (58, 59, 60), THV is missing important nucleic acid metabolism enzymes like thymidylate synthase or dihydrofolate reductase. In addition, THV possesses no thymidine kinase, the gene for which is absent from all betaherpesvirus genomes. An important putative THV gene product is T97, which is homologous to HCMVA UL97 (ganciclovir kinase). This protein is responsible for the ganciclovir effect caused by chemotherapy of HCMV infection (53). THV T98 is homologous to the herpesvirus alkaline exonuclease that plays a role in DNA processing and capsid transport from the nucleus to the cytoplasm (26, 78).

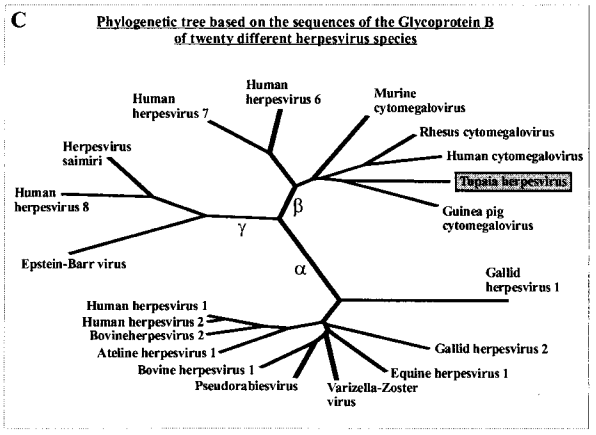
As far as viral glycoproteins are concerned, THV T37, T50, T55 (glycoprotein B), T73, T74 (glycoprotein O), T75 (glycoprotein H), T100 (glycoprotein M), and T115 (glycoprotein L) are homologous to HCMVA and MCMVS glycoproteins. HCMVA UL74, UL75, and UL115 form the gCIII glycoprotein complex (42, 43, 44, 48). The gCIII glycoprotein complex and glycoprotein B are the most important structural surface proteins of HCMV. HCMVA UL55 and UL75 are supposed to start pathways that lead to the activation of cellular transcription factors Sp1 and NFκB by binding to special cellular receptors (93). In the THV genome, 10 potential glycoproteins (t7, t11, t17, t22, t22.1, t22.2, t22.3, t22.5, t22.7, and t22.8) with no homologies to known proteins were identified according to characteristic sites and signatures.

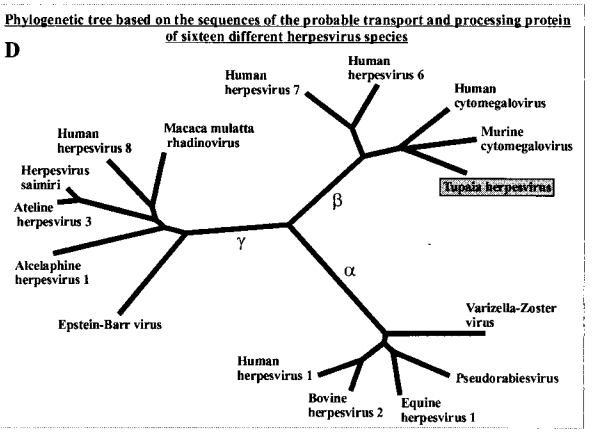
THV T33 and T78 are GCR homologues and are supposed

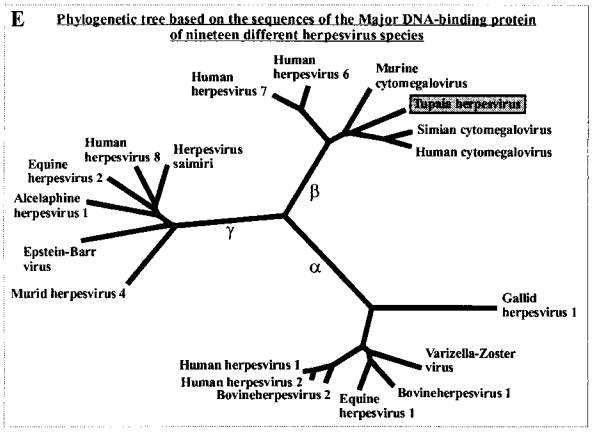
FIG. 6. Eight phylogenetic trees derived by comparison of the DNA polymerase, DNA polymerase accessory protein, glycoprotein B, probable transport and processing protein, major DNA-binding protein, major capsid protein, viral terminase, and uracil DNA-glycosylase amino acid sequences of different herpesvirus species. The three main branches of the trees represent the evolutionary lineages of the herpesvirus subfamilies *Alpha* (α)-, *Beta* (β)-, and *Gammaherpesvirinae* (γ). The sequences of the individual proteins used to construct the phylogenetic trees were taken from the GenBank and SwissPort release 39 databases.

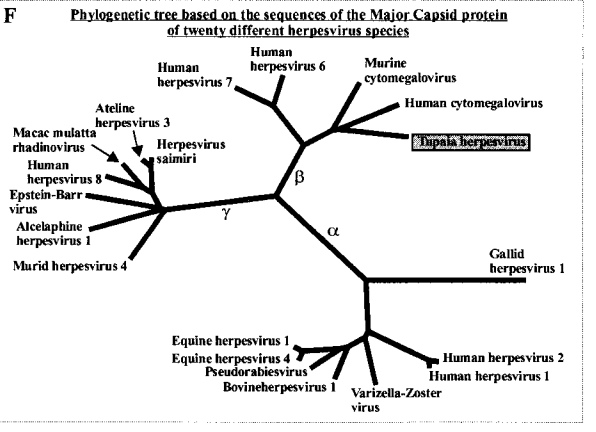

A Phylogenetic tree based on the sequences of the DNA polymerase accessory protein of seventeen different herpesvirus species
Tupaia herpesvirus
Human cytomegalovirus
Murine cytomegalovirus
Human herpesvirus 6
Human herpesvirus 7
Murid herpesvirus 4
Epstein-Barr virus
Equine herpesvirus 2
Herpesvirus saimiri
Alcelaphine herpesvirus 1
Human herpesvirus 8
Equine herpesvirus 1
Equine herpesvirus 4
Pseudo-rabiesvirus
Bovine herpesvirus 1
Human herpesvirus 1
Varizella-Zoster virus
β
γ
α

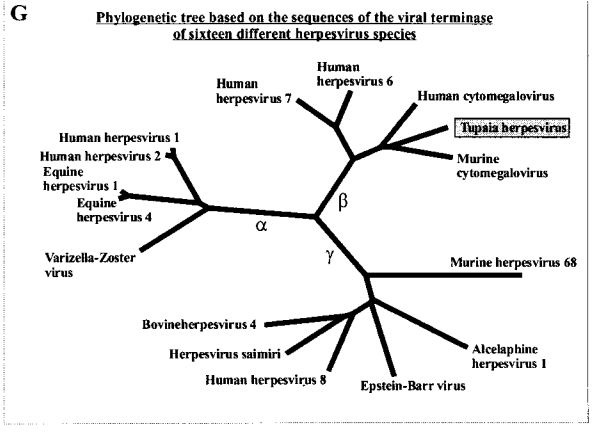

B Phylogenetic tree based on the sequences of the DNA polymerase of nineteen different herpesvirus species
Guinea pig cytomegalovirus
Tupaia herpesvirus
Human herpesvirus 6
Rhesus cytomegalovirus
Human herpesvirus 7
Human cytomegalovirus
Rat cytomegalovirus
Murine cytomegalovirus
Herpesvirus saimiri
Human herpesvirus 8
Equine herpesvirus 2
Alcelaphine herpesvirus 1
Epstein-Barr virus
Gallid herpesvirus 2
Human herpesvirus 1
Human herpesvirus 2
Varizella-Zoster virus
Bovine herpesvirus 2
Equine herpesvirus 1
β
γ
α

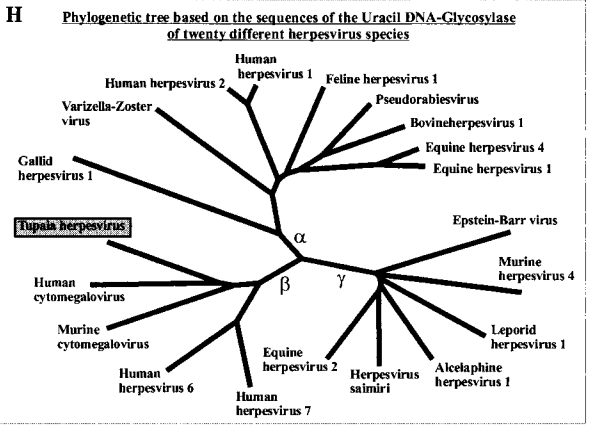

C Phylogenetic tree based on the sequences of the Glycoprotein B of twenty different herpesvirus species
Human herpesvirus 7
Human herpesvirus 6
Murine cytomegalovirus
Rhesus cytomegalovirus
Herpesvirus saimiri
Human cytomegalovirus
Tupaia herpesvirus
Human herpesvirus 8
Guinea pig cytomegalovirus
Epstein-Barr virus
Gallid herpesvirus 1
Human herpesvirus 1
Human herpesvirus 2
Bovineherpesvirus 2
Ateline herpesvirus 1
Gallid herpesvirus 2
Bovine herpesvirus 1
Equine herpesvirus 1
Pseudorabiesvirus
Varizella-Zoster virus
β
γ
α


Phylogenetic tree based on the sequences of the probable transport and processing protein of sixteen different herpesvirus species
D
Human herpesvirus 7
Human herpesvirus 6
Human cytomegalovirus
Macaca mulatta rhadinovirus
Murine cytomegalovirus
Human herpesvirus 8
Herpesvirus saimiri
Tupaia herpesvirus
Ateline herpesvirus 3
Alcelaphine herpesvirus 1
Epstein-Barr virus
Varizella-Zoster virus
Human herpesvirus 1
Pseudorabiesvirus
Bovine herpesvirus 2
Equine herpesvirus 1
β
γ
α


E Phylogenetic tree based on the sequences of the Major DNA-binding protein of nineteen different herpesvirus species
Human herpesvirus 7
Human herpesvirus 6
Murine cytomegalovirus
Tupaia herpesvirus
Simian cytomegalovirus
Human herpesvirus 8
Herpesvirus saimiri
Human cytomegalovirus
Equine herpesvirus 2
Alcelaphine herpesvirus 1
Epstein-Barr virus
Murid herpesvirus 4
Gallid herpesvirus 1
Human herpesvirus 1
Human herpesvirus 2
Bovineherpesvirus 2
Varizella-Zoster virus
Bovineherpesvirus 1
Equine herpesvirus 1
β
γ
α


F Phylogenetic tree based on the sequences of the Major Capsid protein of twenty different herpesvirus species
Human herpesvirus 7
Human herpesvirus 6
Murine cytomegalovirus
Ateline herpesvirus 3
Human cytomegalovirus
Macac mulatta rhadinovirus
Herpesvirus saimiri
Tupaia herpesvirus
Human herpesvirus 8
Epstein-Barr virus
Alcelaphine herpesvirus 1
Murid herpesvirus 4
Gallid herpesvirus 1
Equine herpesvirus 1
Equine herpesvirus 4
Pseudorabiesvirus
Bovineherpesvirus 1
Human herpesvirus 2
Human herpesvirus 1
Varizella-Zoster virus
β
γ
α


G Phylogenetic tree based on the sequences of the viral terminase of sixteen different herpesvirus species
Human herpesvirus 7
Human herpesvirus 6
Human cytomegalovirus
Tupaia herpesvirus
Human herpesvirus 1
Human herpesvirus 2
Murine cytomegalovirus
Equine herpesvirus 1
Equine herpesvirus 4
Varizella-Zoster virus
Murine herpesvirus 68
Bovineherpesvirus 4
Herpesvirus saimiri
Alcelaphine herpesvirus 1
Human herpesvirus 8
Epstein-Barr virus
β
γ
α


H Phylogenetic tree based on the sequences of the Uracil DNA-Glycosylase of twenty different herpesvirus species
Human herpesvirus 1
Human herpesvirus 2
Feline herpesvirus 1
Varizella-Zoster virus
Pseudorabiesvirus
Bovineherpesvirus 1
Gallid herpesvirus 1
Equine herpesvirus 4
Equine herpesvirus 1
Tupaia herpesvirus
Epstein-Barr virus
Murine herpesvirus 4
Human cytomegalovirus
Murine cytomegalovirus
Leporid herpesvirus 1
Equine herpesvirus 2
Herpesvirus saimiri
Alcelaphine herpesvirus 1
Human herpesvirus 6
Human herpesvirus 7
α
β
γ

to be homologous to host proteins. It is presumed that the function of viral GCR homologues is to catch extracellular signals and block the pertinent intracellular pathways. HCMVA UL33 is a CC chemokine receptor (9) and is not essential for growth in cell culture (56). UL33 homologues are conserved only in betaherpesviruses (8, 25). The HHV-6 homologue of THV T78 was found to be a CC chemokine receptor in vitro (67). t90 is not homologous to any known protein but was found to hold a GCR signature. However, further work is necessary to determine if t90 is actually a GCR homologue. US27 and US28 of HCMVA are CC chemokine receptor homologues and also members of the GCR family (9, 95). However, no corresponding homologous proteins could be identified in the THV genome.

The THV genome possesses a number of ORFs that code for proteins with no homologies to known proteins. These genes seem to code for virus species-specific functions in the natural host. t16 was found to possess a potential aminoacyltransfer RNA synthetase class II signature, and t119 holds an ATP-dependent DNA ligase AMP-binding site. However, it has to be verified whether these signatures are actually functional. Interestingly, the locations of the unconserved genes are almost in the same genomic areas within the genomes of the members of the subfamily *Betaherpesvirinae*. The majority of these ORFs code for glycoproteins and are members of gene families. HCMVA UL24, UL43, UL83, UL116, UL118, and UL121 have corresponding homologues in MCMVS but not in THV. The latter five seem to have functions that are specific to cytomegaloviruses.

The biological and genomic properties of THV correspond to the criteria prepared for the classification of herpesviruses in the subfamily *Betaherpesvirinae* (73). THV is evolutionarily placed within the group of mammalian cytomegaloviruses. There is good reason to suppose that the separation of the three evolutionary lineages which lead to THV and the primate and the rodent cytomegaloviruses has taken place in a very short evolutionary period. This assumption is in accordance with the phylogenetic tree of the hosts of these virus species and confirms the accepted hypothesis that herpesviruses follow the development of their hosts, a process known as coevolution (47, 61, 62, 63). The identification of the genetic structure of the viral ancestor of these three herpesviruses is very difficult. It is certain that the homologous genes of THV, HCMVA, and MCMVS were also present in the ancestral genome. However, it is not clear why the gene arrangements of THV and MCMVS correspond to two different HCMVA genome isomers. A simple explanation would be that the viral ancestor had repetitive DNA elements similar to those of HCMVA and the ability to form genome isomers. However, one can assume that the repetitive DNA elements disappeared in the MCMVS and THV phylogenetic lineages due to distinct evolutionary pressures. In view of the high degree of relatedness of THV to the mammalian cytomegaloviruses, THV can be considered a model system for the investigation of HCMV infection and pathogenesis.

ACKNOWLEDGMENTS

We thank the Deutsche Forschungsgemeinschaft, project Da-142/10-1-5, for providing the automatic DNA sequencing equipment (373 "Extended" DNA sequencer; Perkin-Elmer Corporation, Applied Biosystems).

We thank Michaela Handermann and Nurith J. Jakob for critical comments and helpful discussion.

REFERENCES

1. **Al-Barazi, H. O., and A. M. Colberg-Poley.** 1996. The human cytomegalovirus UL37 immediate-early regulatory protein is an integral membrane N-glycoprotein which traffics through the endoplasmic reticulum and Golgi apparatus. J. Virol. **70:**7198–7208.
2. **Albrecht, M., G. Darai, and R. M. Flügel.** 1985. Analysis of the genomic termini of tupaia herpesvirus DNA by restriction mapping and nucleotide sequencing. J. Virol. **56:**466–474.
3. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
4. **Baer, R., A. T. Bankier, M. D. Biggin, P. L. Deininger, P. J. Farrell, T. J. Gibson, G. Hatfull, G. S. Hudson, S. C. Satchwell, C. Seguin, P. S. Tuffnell, and B. G. Barrell.** 1984. DNA sequence and expression of the B95–8 Epstein-Barr virus genome. Nature **310:**207–211.
5. **Bahr, U., C. Springfeld, C. A. Tidona, and G. Darai.** 1999. Structural organization of a conserved gene cluster of Tupaia herpesvirus encoding the DNA polymerase, glycoprotein B, a probable processing and transport protein, and the major DNA binding protein. Virus Res. **60:**123–136.
6. **Battista, M. C., G. Bergamini, M. C. Boccuni, F. Campanini, A. Ripalti, and M. P. Landini.** 1999. Expression and characterization of a novel structural protein of human cytomegalovirus, pUL25. J. Virol. **73:**3800–3809.
7. **Baum, E. Z., G. A. Bebernitz, J. D. Hulmes, V. P. Muzithras, T. R. Jones, and Y. Gluzman.** 1993. Expression and analysis of the human cytomegalovirus UL80-encoded protease: identification of autoproteolytic sites. J. Virol. **67:**497–506.
8. **Besieger, P. S., C. Vink, J. G. Van Dam, G. Grauls, S. J. Vanherle, and C. A. Bruggeman.** 1998. The R33 G protein-coupled receptor gene of rat cytomegalovirus plays an essential role in the pathogenesis of viral infection. J. Virol. **72:**2352–2363.
9. **Bodaghi, B., T. R. Jones, D. Zipeto, C. Vita, L. Sun, L. Laurent, F. Arenzana-Seisdedos, J. L. Virelizier, and S. Michelson.** 1998. Chemokine sequestration by viral chemoreceptors as a novel viral escape strategy: withdrawal of chemokines from the environment of cytomegalovirus-infected cells. J. Exp. Med. **188:**855–866.
10. **Bucher, P., and A. Bairoch.** 1994. A generalized profile syntax for biomolecular sequence motifs and its function in automatic sequence interpretation, p. 53–61. R. Altman, D. Brutlag, P. Karp, R. Lathrop, and D. Searls (ed.), *In* ISMB-94. Proceedings of the 2nd International Conference on Intelligent Systems for Molecular Biology. AAAI Press, Menlo Park, Calif.
11. **Burck, P. J., D. H. Berg, T. P. Luk, L. M. Sassmannshausen, M. Walkulchik, D. P. Smith, H. M. Hsiung, G. W. Becker, W. Gibson, and E. C. Villarreal.** 1994. Human cytomegalovirus maturational proteinase: expression in *Escherichia coli*, purification, and enzymatic characterization by using peptide substrate mimics of natural cleavage sites. J. Virol. **68:**2937–2946.
12. **Cardin, R. D., G. B. Abenes, C. A. Stoddart, and E. S. Mocarski.** 1995. Murine cytomegalovirus IE2, an activator of gene expression, is dispensible for growth and latency in mice. Virology **209:**236–241.
13. **Chee, M. S., A. T. Bankier, S. Beck, R. Bohni, C. M. Brown, R. Cerny, 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 the sequence of human cytomegalovirus strain AD169. Curr. Top. Microbiol. Immunol. **154:**125–169.
14. **Colberg-Poley, A. M.** 1996. Functional roles of immediate early proteins encoded by the human cytomegalovirus UL36–38, UL115–119, TRS1/IRS1 and US3 loci. Intervirology **39:**350–360.
15. **Colberg-Poley, A. M., L. Huang, V. E. Soltero, A. C. Iskenderian, R. F. Schumacher, and D. G. Anders.** 1998. The acidic domain of pUL37 × 1 and gpUL37 plays a key role in transactivation of HCMV DNA replication gene promoter constructions. Virology **246:**400–408.
16. **Cranmer, L. D., C. Clark, and D. H. Spector.** 1994. Cloning, characterization, and expression of the murine cytomegalovirus homologue of the human cytomegalovirus 28-kDa matrix phosphoprotein (UL99). Virology **205:**417–429.
17. **Cranmer, L. D., C. L. Clark, C. S. Morello, H. E. Farrell, W. D. Rawlinson, and D. H. Spector.** 1996. Identification, analysis, and evolutionary relationships of the putative murine cytomegalovirus homologs of the human cytomegalovirus UL82 (pp71) and UL83 (pp65) matrix phosphoproteins. J. Virol. **70:**7929–7939.
18. **Dallas, P. B., P. A. Lyons, J. B. Hudson, A. A. Scalzo, and G. R. Shellam.** 1994. Identification and characterization of a murine cytomegalovirus gene with homology to the UL25 open reading frame of human cytomegalovirus. Virology **200:**643–650.
19. **Darai, G., B. Matz, C. H. Schroeder, R. M. Flügel, U. Berger, K. Munk, and H. Gelderblom.** 1979. Characterization of a tree shrew herpesvirus isolated from a lymphosarcoma. J. Gen. Virol. **43:**541–551.
20. **Darai, G., L. Zöller, B. Matz, R. M. Flügel, W. Hofmann, and H. Gelder-**

blom. 1980. Herpesvirus tupaia: isolation, characterization and oncogenicity, p. 141–148. *In* B. Lapin and S. Yohn (ed.), Advances in leucaemia research, 1979. USSR Academy of Medical Science, IEPT, Moscow, Russia.
21. **Darai, G., R. M. Flügel, B. Matz, and H. Delius.** 1981. DNA of tupaia herpesviruses, p. 345–361. *In* Y. Becker (ed.), Herpes virus DNA. Martinus Nijhoff, The Hague, The Netherlands.
22. **Darai, G., H. G. Koch, R. M. Flügel, and H. Gelderblom.** 1983. Tree shrew (Tupaia) herpesviruses, p. 39–51. *In* M. Bonneau and W. Hennessen (ed.), Herpesvirus of man and animal. Developments in biological standardization. Karger, Basel, Switzerland.
23. **Darai, G., A. Rosen, J. Scholz, and H. Gelderblom.** 1983. Induction of generalized and lethal herpesvirus infection in the tree shrew by intrahepatic transfection of herpes simplex virus DNA. J. Virol. Methods **7:**305–314.
24. **Darai, G., and H. G. Koch.** 1984. Tree shrew herpesvirus: pathogenicity and latency, p. 1–102. *In* G. Wittmann, R. M. Gaskell, and R. J. Rziha (ed.), Latent herpesvirus infections in veterinary medicine. Martinus Nijhoff, The Hague, The Netherlands.
25. **Davis-Poynter, N. J., D. M. Lynch, H. Vally, G. R. Shellam, W. D. Rawlinson, B. G. Barrell, and H. E. Farrell.** 1997. Identification and characterization of a G protein-coupled receptor homolog encoded by murine cytomegalovirus. J. Virol. **71:**1521–1529.
26. **Gao, M., B. J. Robertson, P. J. McCann, D. R. O'Boyle, S. K. Weller, W. W. Newcomb, J. C. Brown, and S. P. Weinheimer.** 1998. Functional conservations of the alkaline nuclease of herpes simplex type 1 and human cytomegalovirus. Virology **249:**460–470.
27. **Gebert, S., S. Schmolke, G. Sorg, S. Floss, B. Plachter, and T. Stamminger.** 1997. The UL84 protein of human cytomegalovirus acts as a transdominant inhibitor of immediate-early-mediated transactivation that is able to prevent viral replication. J. Virol. **71:**7048–7060.
28. **Gibson, W., K. S. Clopper, W. J. Britt, and M. K. Baxter.** 1996. Human cytomegalovirus (HCMV) smallest capsid protein identified as product of short open reading frame located between HCMV UL48 and UL49. J. Virol. **70:**5680–5683.
29. **Gibson, W., M. K. Baxter, and K. S. Clopper.** 1996. Cytomegalovirus "missing" capsid protein identified as heat-aggregable product of human cytomegalovirus UL46. J. Virol. **70:**7454–7461.
30. **Goldmacher, V. S., L. M. Bartle, A. Skaletskaya, C. A. Dionne, N. L. Kedersha, C. A. Vater, J. W. Han, R. J. Lutz, S. Watanabe, E. D. McFarland, E. D. Kieff, E. S. Mocarski, and T. Chittenden.** 1999. A cytomegalovirus-encoded mitochondria-localized inhibitor of apoptosis structurally unrelated to Bcl-2. Proc. Natl. Acad. Sci. USA **96:**12536–12541.
31. **Gompels, U. A., J. Nicholas, G. Lawrence, M. Jones, B. J. Thomson, M. E. D. Martin, S. Efstathiou, M. Craxton, and H. A. Macaulay.** 1995. The DNA sequence of human herpesvirus-6: structure, coding content and genome evolution. Virology **209:**29–51.
32. **Greenaway, P. J., and G. W. G. Wilkinson.** 1987. Nucleotide sequence of the most abundantly transcribed early gene of human cytomegalovirus strain AD169. Virus Res. **7:**17–31.
33. **Greijer, A. E., J. M. van de Crommert, S. J. Stevens, and J. M. Middeldorp.** 1999. Molecular fine-specificity analysis of antibody responses to human cytomegalovirus and design of novel synthetic-peptide-based serodiagnostic assays. J. Clin. Microbiol. **37:**179–188.
34. **Greis, K. D., W. Gibson, and G. W. Hart.** 1994. Site-specific glycosylation of the human cytomegalovirus tegument basic phosphoprotein (UL32) at serine 921 and serine 952. J. Virol. **68:**8339–8349.
35. **Hanson, L. K., B. L. Dalton, Z. Karabekian, H. E. Farell, W. D. Rawlinson, R. M. Stenberg, and A. E. Campbell.** 1999. Transcriptional analysis of the murine cytomegalovirus HindIII-I region: identification of a novel immediate-early gene region. Virology **260:**156–164.
36. **Hayashi, M. L., C. Blankenship, and T. Shenk.** 2000. Human cytomegalovirus UL69 protein is required for efficient accumulation of infected cells in the G1 phase of the cell cycle. Proc. Natl. Acad. Sci. USA **97:**2692–2696.
37. **Hensel, G., H. Meyer, S. Gartner, G. Brand, and H. F. Kern.** 1995. Nuclear localization of the human cytomegalovirus tegument protein pp150 (ppUL32). J. Gen. Virol. **76:**1591–1601.
38. **Hensel, G. M., H. H. Meyer, I. Buchmann, D. Pommerehne, S. Schmolke, B. Plachter, K. Radsak, and H. F. Kern.** 1996. Intracellular localization and expression of the human cytomegalovirus matrix phosphoprotein pp71 (ppUL82): evidence for its translocation into the nucleus. J. Gen. Virol. **7:**3087–3097.
39. **Higgins, D. G., and P. M. Sharp.** 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene **73:**237–244.
40. **Hofmann, K., P. Bucher, L. Falquet, and A. Bairoch.** 1999. The PROSITE database, its status in 1999. Nucleic Acids Res. **27:**215–219.
41. **Hofmann, W., P. Moller, A. Schwaier, R. M. Flugel, L. Zoller, and G. Darai.** 1981. Malignant tumours in tupaia (tree shrew). J. Med. Primatol. **10:**155–163.
42. **Huber, M. T., and T. Compton.** 1997. Characterization of a novel third member of the human cytomegalovirus glycoprotein H-glycoprotein L complex. J. Virol. **71:**5391–5398.
43. **Huber, M. T., and T. Compton.** 1998. The human cytomegalovirus UL74 gene encodes the third component of the glycoprotein H-glycoprotein L-containing envelope complex. J. Virol. **72:**8191–8197.
44. **Huber, M. T., and T. Compton.** 1999. Intracellular formation and processing of the heterotrimeric gH-gL-gO (gCIII) glycoprotein envelope complex of human cytomegalovirus. J. Virol. **73:**3886–3892.
45. **Hutchinson, N. I., and M. J. Tocci.** 1986. Characterization of a major early gene from the human cytomegalovirus long inverted repeat; predicted amino acid sequence of a 30-kDa protein encoded by the 1.2kb mRNA. Virology **155:**172–182.
46. **Iskenderian, A. C., L. Huang, A. Reilly, R. M. Stenberg, and D. G. Anders.** 1996. Four of eleven loci required for transient complementation of human cytomegalovirus DNA replication cooperate to active expression of replication genes. J. Virol. **70:**383–392.
47. **Karlin, S., E. S. Mocarski, and G. A. Schachtel.** 1994. Molecular evolution of herpesviruses: genomic and protein sequence comparisons. J. Virol. **68:** 1886–1902.
48. **Kaye, J. F., U. A. Gompels, and A. C. Minson.** 1992. Glycoprotein H of human cytomegalovirus (HCMV) forms a stable complex with the HCMV UL115 gene product. J. Gen. Virol. **73:**2693–2698.
49. **Koch, H. G., H. Delius, B. Matz, R. M. Flugel, J. Clarke, and G. Darai.** 1985. Molecular cloning and physical mapping of the tupaia herpesvirus genome. J. Virol. **55:**86–95.
50. **Koch, H. G., R. M. Flugel, and G. Darai.** 1986. Concatemeric forms of intracellular tupaia herpesvirus DNA. Virology **151:**211–221.
51. **Krosky, P. M., M. R. Underwood, S. R. Turk, K. W. H. Feng, R. K. Jain, R. G. Ptak, A. C. Westerman, K. K. Biron, L. B. Townsend, and J. C. Drach.** 1998. Resistance of human cytomegalovirus to benzimidazole ribonucleosides maps to two open reading frames: UL89 and UL56. J. Virol. **72:**4721–4728.
52. **Kumar, S., and S. B. Hedges.** 1998. A molecular timescale for vertebrate evolution. Nature **392:**917–920.
53. **Littler, E., A. D. Stuart, and M. S. Chee.** 1992. Human cytomegalovirus UL97 open reading frame encodes a protein that phosphorylates the antiviral nucleoside analogue ganciclovir. Nature **358:**160–162.
54. **Liu, F., and B. Roizman.** 1992. Differentiation of multiple domains in the herpes simplex virus 1 protease encoded by the UL26 gene. Proc. Natl. Acad. Sci. USA **89:**2076–2080.
55. **Lu, M., and T. Shenk.** 1999. Human cytomegalovirus UL69 protein induces cells to accumulate in G1 phase of the cell cycle. J. Virol. **73:**676–683.
56. **Margulies, B. J., H. Browne, and W. Gibson.** 1996. Identification of the human cytomegalovirus G protein-coupled receptor homologue encoded by UL33 in infected cells and enveloped virus particles. Virology **225:**111–125.
57. **McCombs, R. M., J. P. Brunschwig, R. Mircovic, and M. Benyesh-Melnick.** 1971. Electron microscopic characterization of a herpes-like virus isolated from tree shrews. Virology **45:**816–820.
58. **McGeoch, D. J., A. Dolan, S. Donald, and F. J. Rixon.** 1985. Sequence determination and genetic content of the short unique region in the genome of herpes simplex virus type I. J. Mol. Biol. **181:**1–13.
59. **McGeoch, D. J., A. Dolan, S. Donald, and D. H. Brauer.** 1986. Complete DNA sequence of the short repeat region in the genome of herpes simplex virus type 1. Nucleic Acids Res. **14:**1727–1745.
60. **McGeoch, D. J., M. A. Dalrymple, A. J. Davison, A. Dolan, M. C. Frame, D. McNab, L. J. Perry, J. E. Scott, and P. Taylor.** 1988. The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. J. Gen. Virol. **69:**1531–1574.
61. **McGeoch, D. J.** 1989. The genomes of the human herpesviruses: contents, relationship and evolution. Annu. Rev. Microbiol. **43:**235–265.
62. **McGeoch, D. J., and S. Cook.** 1994. Molecular phylogeny of the alphaherpesvirinae subfamily and a proposed evolutionary timescale. J. Mol. Biol. **238:**9–22.
63. **McGeoch, D. J., S. Cook, A. Dolan, F. E. Jamieson, and E. A. Telford.** 1995. Molecular phylogeny and evolutionary timescale for the family of mammalian herpesviruses. J. Mol. Biol. **247:**443–458.
64. **McLauchlan, J.** 1997. The abundance of the herpes simplex virus type 1 UL37 tegument protein in virus particles is closely controlled. J. Gen. Virol. **78:**189–194.
65. **Messerle, M., G. M. Keil, and U. H. Koszinowski.** 1991. Structure and expression of murine cytomegalovirus immediate-early gene 2. J. Virol. **65:** 1638–1643.
66. **Meyer, H. H., A. Ripalti, M. P. Landini, K. Radsak, H. F. Kern, and G. M. Hensel.** 1997. Human cytomegalovirus late-phase maturation is blocked by stably expressed UL32 antisense mRNA in astrocytoma cells. J. Gen. Virol. **78:**2621–2631.
67. **Milne, R. S. B., C. Mattick, L. Nicholson, P. Devaray, A. Alcami, and U. Gompels.** 2000. RANTES binding and down-regulation by a novel human herpesvirus-6β chemokine receptor. J. Immunol. **164:**2396–2404.
68. **Mirkovic, R., W. R. Voss, and M. Beyesh-Melnick.** 1970. Characterization of a new herpes-type virus indigenous for tree shrews, p. 181–189. Proceedings of the 10th International Congress of Microbiology, Mexico City, Mexico.
69. **Novacek, M. J.** 1992. Mammalian phylogeny: shaking the tree. Nature **356:** 121–125.
70. **Pari, G. S., and D. G. Anders.** 1993. Eleven loci encoding *trans*-acting factors are required for transient complementation of human cytomegalovirus *ori-*

Lyt-dependent DNA replication. J. Virol. **67:**6979–6988.
71. **Rawlinson, W. D., H. E. Farrell, and B. G. Barrell.** 1996. Analysis of the complete DNA sequence of murine cytomegalovirus. J. Virol **70:**8833–8849.
72. **Roizman, B., L. E. Carmichael, F. Deinhardt, G. de-The, A. J. Nahmias, W. Plowright, F. Rapp, P. Sheldrick, M. Takahashi, and K. Wolf.** 1981. Herpesviridae: definition, provisional nomenclature, and taxonomy. Intervirology **16:**201–217.
73. **Roizman, B.** 1996. Herpesviridae, p. 2221–2230. *In* B. N. Fields, D. M. Knipe, P. M. Howley, et al. (ed.), Fields virology, 3rd ed. Lippincott-Raven Publishers, Philadelphia, Pa.
74. **Rösen-Wolff, A., T. Ben-Hur, Y. Becker, and G. Darai.** 1988. Comparative analysis of the transcripts mapped in the BamHI DNA fragment B of avirulent HSV-1 HFEM, virulent HSV-1 F, and their intratypic recombinant viruses. Virus Res. **10:**315–324.
75. **Sardana, V. V., J. A. Wolfgang, C. A. Veloski, W. J. Long, K. LeGrow, B. Wolanski, E. A. Emini, and R. L. LaFemina.** 1994. Peptide substrate cleavage specificity of the human cytomegalovirus protease. J. Biol. Chem. **269:** 14337–14340.
76. **Sarisky, R. T., and G. S. Hayward.** 1996. Evidence that the UL84 gene product of human cytomegalovirus is essential for promoting *ori*Lyt-dependent DNA replication and formation of replication compartments in cotransfection assays. J. Virol. **70:**7398–7413.
77. **Schmolke, S., H. F. Kern, P. Drescher, G. Jahn, and B. Plachter.** 1995. The dominant phosphoprotein pp65 (UL83) of human cytomegalovirus is dispensable for growth in cell culture. J. Virol. **69:**5959–5968.
78. **Sheaffer, A. K., S. P. Weinheimer, and D. J. Tenney.** 1997. The human cytomegalovirus UL98 gene encodes the conserved herpesvirus alkaline nuclease. J. Gen. Virol. **78:**2953–2961.
79. **Smith, J. A., and G. S. Pari.** 1995. Expression of human cytomegalovirus UL36 and UL37 genes is required for viral DNA replication. J. Virol. **69:**1925–1931.
80. **Smith, J. A., and G. S. Pari.** 1995. Human cytomegalovirus UL102 gene. J. Virol. **69:**1734–1740.
81. **Smith, J. A., S. Jairath, J. J. Crute, and G. S. Pari.** 1996. Characterization of the human cytomegalovirus UL105 gene and identification of the putative helicase protein. Virology **220:**251–255.
82. **Springfeld, C., C. A. Tidona, R. Kehm, U. Bahr, and G. Darai.** 1998. Identification and characterization of the tupaia herpesvirus DNA polymerase gene. J. Gen. Virol. **79:**3049–3053.
83. **Stenberg, R. M., D. R. Thomsen, and M. F. Stinski.** 1984. Structural analysis of the major immediate early gene of human cytomegalovirus. J. Virol. **49:**190–191.
84. **Stenberg, R. M., P. R. Witte, and M. F. Stinski.** 1985. Multiple spliced and unspliced transcripts from human cytomegalovirus immediate-early region 2 and evidence for a common initiation site within immediate-early region 1. J. Virol. **56:**665–675.
85. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:**4876–4882.
86. **Tidona, C. A., and G. Darai.** 1997. The complete DNA sequence of lymphocystis disease virus. Virology **230:**207–216.
87. **Vink, C., E. Beuken, and C. A. Bruggeman.** 2000. Complete DNA sequence of the rat cytomegalovirus genome. J. Virol. **74:**7656–7665.
88. **Weir, J. P.** 1998. Genomic organization and evolution of the human herpesviruses. Virus Genes **16:**85–93.
89. **Wing, B. A., G. C. Lee, and E.-S. Huang.** 1996. The human cytomegalovirus UL94 open reading frame encodes a conserved herpesvirus capsid/tegument-associated virion protein that is expressed with true late kinetics. J. Virol. **70:**3339–3345.
90. **Winkler, M., S. A. Rice, and T. Stamminger.** 1994. UL69 of human cytomegalovirus, an open reading frame with homology to ICP27 of herpes simplex virus, encodes a transactivator of gene expression. J. Virol. **68:**3943–3954.
91. **Winkler, M., and T. Stamminger.** 1996. A specific subform of the human cytomegalovirus transactivator protein pUL69 is contained within the tegument of virus particles. J. Virol. **70:**8984–8987.
92. **Wood, L. J., M. K. Baxter, S. M. Plafker, and W. Gibson.** 1997. Human cytomegalovirus capsid assembly protein precursor (pUL80.5) interacts with itself and with the major capsid protein (pUL86) through two different domains. J. Virol. **71:**179–190.
93. **Yurochko, A. D., E.-S. Hwang, L. Rasmussen, S. Keay, L. Pereira, and E.-S. Huang.** 1997. The human cytomegalovirus UL55 (gB) and UL75 (gH) glycoprotein ligands initiate the rapid activation of Sp1 and NF-κB during infection. J. Virol. **71:**5051–5059.
94. **Zini, N., M. C. Battista, S. Santi, M. Riccio, G. Bergamini, M. P. Landini, and N. M. Maraldi.** 1999. The novel structural protein of human cytomegalovirus, pUL25, is localized in the viral tegument. J. Virol. **73:**6073–6075.
95. **Zipeto, D., B. Bodaghi, L. Laurent, J. L. Virelizier, and S. Michelson.** 1999. Kinetics of transcription of human cytomegalovirus chemokine receptor US28 in different cell types. J. Gen. Virol. **80:**543–547.