JVI Journals.ASM.org

Comparative Genome Analysis of Four Elephant Endotheliotropic Herpesviruses, EEHV3, EEHV4, EEHV5, and EEHV6, from Cases of Hemorrhagic Disease or Viremia

Jian-Chao Zong,[a] Erin M. Latimer,[b] Simon Y. Long,[a] Laura K. Richman,[b] Sarah Y. Heaggans,[a] Gary S. Hayward[a]

Viral Oncology Program, The Sidney Kimmel Comprehensive Cancer Center, The Johns Hopkins School of Medicine, Baltimore, Maryland, USA[a]; National Elephant Herpesvirus Laboratory, Pathology Department, Smithsonian's National Zoo, Washington, DC, USA[b]

ABSTRACT

The genomes of three types of novel endotheliotropic herpesviruses (elephant endotheliotropic herpesvirus 1A [EEHV1A], EEHV1B, and EEHV2) associated with lethal hemorrhagic disease in Asian elephants have been previously well characterized and assigned to a new *Proboscivirus* genus. Here we have generated 112 kb of DNA sequence data from segments of four more types of EEHV by direct targeted PCR from blood samples or necropsy tissue samples from six viremic elephants. Comparative phylogenetic analysis of nearly 30 protein-encoding genes of EEHV5 and EEHV6 show that they diverge uniformly by nearly 20% from their closest relatives, EEHV2 and EEHV1A, respectively, and are likely to have similar overall gene content and genome organization. In contrast, seven EEHV3 and EEHV4 genes analyzed differ from those of all other EEHVs by 37% and have a G+C content of 63% compared to just 42% for the others. Three strains of EEHV5 analyzed clustered into two partially chimeric subgroups EEHV5A and EEHV5B that diverge by 19% within three small noncontiguous segments totaling 6.2 kb. We conclude that all six EEHV types should be designated as independent species within a proposed new fourth *Deltaherpesvirinae* subfamily of mammalian herpesviruses. These virus types likely initially diverged close to 100 million years ago when the ancestors of modern elephants split from all other placental mammals and then evolved into two major branches with high- or low-G+C content about 35 million years ago. Later additional branching events subsequently generated three paired sister taxon lineages of which EEHV1 plus EEHV6, EEHV5 plus EEHV2, and EEHV4 plus EEHV3 may represent Asian and African elephant versions, respectively.

IMPORTANCE

One of the factors threatening the long-term survival of endangered Asian elephants in both wild range countries and in captive breeding populations in zoos is a highly lethal hemorrhagic herpesvirus disease that has killed at least 70 young Asian elephants worldwide. The genomes of the first three types of EEHVs (or proboscivirus) identified have been partially characterized in the preceding accompanying paper (L. K. Richman, J.-C. Zong, E. M. Latimer, J. Lock, R. C. Fleischer, S. Y. Heaggans, and G. S. Hayward, J. Virol. 88:13523–13546, 2014, http://dx.doi.org/10.1128/JVI.01673-14). Here we have used PCR DNA sequence analysis from multiple segments of DNA amplified directly from blood or necropsy tissue samples of six more selected cases of hemorrhagic disease to partially characterize four other types of EEHVs from either Asian or African elephants. We propose that all six types and two chimeric subtypes of EEHV belong to multiple lineages of both AT-rich and GC-rich branches within a new subfamily to be named the *Deltaherpesvirinae*, which evolved separately from all other mammalian herpesviruses about100 million years ago.

Six distinct groups of a family of novel elephant endotheliotropic herpesviruses (EEHVs) that are associated with a highly lethal systemic hemorrhagic disease occurring predominantly in young Asian elephants have been described (1–5). Their identification was based on characteristic differences in the DNA sequences of small diagnostic PCR products from the terminase (TER) and DNA polymerase (POL) genes. Four of these viruses, EEHV1, EEHV3, EEHV4, and EEHV5, were first detected in blood or necropsy tissue from cases of acute viremic disease in Asian elephants, whereas just three known cases of this disease in African elephants involved two other related viruses, EEHV2 and EEHV6 (1, 4). The vast majority (over 90%) of the 47 PCR sequence-confirmed cases of lethal hemorrhagic disease in captive zoo elephants in Europe and North America, as well as nine Asian calf deaths in India (6), have been caused by a large variety of distinctive strains of two partially chimeric subgroups referred to as EEHV1A and EEHV1B. However, six other elephant calf deaths and several cases of mild hemorrhagic disease have instead been associated with either EEHV2 (1), EEHV3 (3), EEHV4 (3, 7), EEHV5 (4, 8), or EEHV6 (4). In addition, a large family of five distinctive and highly diverged gammaherpesviruses (elephant gammaherpesvirus 1 [EGHV1] to EGHV5) has also been detected mostly in eye and genital swabs from adult Asian and African

Received 11 June 2014 **Accepted** 8 August 2014
Published ahead of print 17 September 2014
Editor: L. Hutt-Fletcher
Address correspondence to Gary S. Hayward, ghayward@jhmi.edu.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/JVI.01675-14.

doi:10.1128/JVI.01675-14

TABLE 1 Summary of six EEHV-positive elephant cases evaluated in these studies

Case	Virus type	Strain	Elephant name[a]	Host species, sex, and age[b]	Location	Yr	Pathology[c] (reference)	DNA source[d]	Sequenced DNA (bp)
1	EEHV3	NAP27	Hansa	EM, F, 7y	Seattle, WA	2007	Fatality (3)	Necropsy tissue sample	4,117
2	EEHV4	NAP22	NA	EM, F, 5y	Oklahoma	2004	Fatality (3)	Necropsy tissue sample	5,743
3	EEHV5A	NAP28	NA	EM, F, 69y	Washington, DC	2007	Routine (4)	WB sample	14,615
4	EEHV5A	NAP50	Methai2	EM, F, 40y	Texas	2011	Letharg. (17)	WB sample	26,090
5	EEHV5B	NAP58	Tucker3	EM, M, 8y	Texas	2011	Asympt. (17)	WB, TW samples	29,347
6	EEHV6[e]	NAP35	NA	LA, F, 15m	Arkansas	2009	Sympt. (4)	WB sample	31,828
Total									111,749

[a] NA, not available.
[b] The host animal species (*Elephas maximus* [EM] or *Loxodonta africana* [LA]), sex (female [F] or male [M]), and age (in months [m] or years [y]) is shown.
[c] Letharg., lethargic; Asympt, asymptomatic; Sympt, symptomatic.
[d] WB, whole blood; TW, trunk wash fluid.
[e] Survived after FCV treatment.

elephants (4, 9), but the latter have not yet been associated with any disease conditions.

None of the EEHVs have been successfully grown or propagated in cell culture, and therefore, all genetic characterization has, of necessity, been carried out directly on total DNA from viremic blood, trunk wash fluid, or necropsy tissue samples. Sanger PCR-based DNA sequence data for between 20 to 70 kb each from seven representative strains of either EEHV1A and EEHV1B, as well as for 59 kb from a single prototype strain of EEHV2, have been generated and described in the preceding accompanying paper (10), adding to two earlier less extensive studies (11, 12). Furthermore, the complete intact 180-kb genomes of three more EEHV1 strains were determined recently by *de novo* next-generation sequencing of necropsy tissue samples (13, 14). The genomes of EEHV1A, EEHV1B, and EEHV2 together form a very distinctive family of herpesviruses with unique genomic organization characteristics. They encompass a total of about 115 genes, including encoding 50 predicted novel proteins, as well as highly diverged orthologues of most common core herpesvirus proteins, but the latter usually branch between those of the classical *Betaherpesvirinae* and *Gammaherpesvirinae* in phylogenetic analyses and are not closely associated with either. Because of these unique characteristics, the EEHVs have been officially classified as members of a new *Proboscivirus* genus that was originally placed within the *Betaherpesvirinae* subfamily (15, 16). However, we have proposed that their deeply diverged phylogenetic position and overall gene organization might make it more appropriate to instead assign EEHV1A, EEHV1B, and EEHV2 to a separate subfamily of mammalian herpeviruses that would logically be designated the *Deltaherpesvirinae* (10, 13).

To understand more about the overall genetic organization, gene content, and evolutionary origin of four other types of EEHVs identified from both Asian and African elephants, we undertook here to further characterize and compare multiple segments of the primary genomic DNA sequence of the prototype strains of EEHV3, EEHV4, EEHV5, and EEHV6. Again, this was determined by targeted Sanger PCR sequencing of more than 112 kb of DNA obtained directly from necropsy tissue samples derived from six selected elephants in the United States that either died of hemorrhagic disease or that survived after PCR-confirmed EEHV-associated viremia. Both EEHV3 and EEHV4 were first reported by Garner et al. (3) at high levels in blood from two young Asian elephants that died suddenly of hemorrhagic disease in 2007 and 2002, respectively. The first example of EEHV5 was detected in blood from an elderly Asian zoo elephant in 2008, and the first example of EEHV6 was in blood from a 1-year-old African elephant calf in 2009, both of which had low-grade symptoms and recovered (4). Mild or asymptomatic infections with EEHV5 were also observed to occur sequentially over a 6-month period in 2010 in all seven members of an Asian elephant herd in Texas when monitoring both blood and trunk wash fluid secretions (17). Two distinct subtypes of EEHV5 referred to here as EEHV5A and EEHV5B each infecting multiple adults and calves proved to be involved in that episode. Finally, a young Asian elephant that died of hemorrhagic disease at a European zoo in 2012 represents the first known fatality associated with EEHV5 (8), and a second lethal case of EEHV4 disease was reported recently in Thailand (7). Initial sequence analysis of just the small POL (480-bp) and TER (340-bp) loci from each of these four new types (3, 4) indicated that EEHV3 and EEHV4 might form a distinctive GC-rich branch of the *Proboscivirus* genus, whereas EEHV1, EEHV2, EEHV5, and EEHV6 fall into an AT-rich branch. The goal of this work was to obtain a much more robust analysis of the phylogenetic and evolutionary relationships among this diverse group of elephant herpesviruses.

MATERIALS AND METHODS

Sources of EEHV genome DNA. The six cases of viremic EEHV in captive-born North American juvenile elephants for which we have carried out detailed DNA sequence analysis here are summarized in Table 1. The sources of five other reference EEHV genomes used extensively for comparisons, namely, EEHV1A(Kala,NAP18), EEHV1A(Kumari,NAP11), EEHV1B(Kiba, NAP14), EEHV1B(Haji, NAP19), and EEHV2(Kijana, NAP12), were described in the preceding accompanying paper (10), and the source of the complete genome of EEHV1A(Kimba, NAP23) was given by Ling et al. (13).

DNA extraction procedures. Total cell DNA was extracted by standard procedures described previously (17–19) from frozen diseased necropsy tissue samples (stored at −80C) or from diagnostic whole blood or trunk wash fluid samples that had been forwarded for analysis to the National Elephant Herpesvirus Laboratory at the Smithsonian's National Zoo in Washington, DC.

G-Phi and PCR amplification, DNA sequencing, and phylogenetic analysis. The G-Phi and PCR amplification, DNA sequencing, and phylogenetic analysis procedures were all described in the preceding accompanying paper by Richman et al. (10) including the MEGA5 (20) and SimPlot programs (21).

PCR amplification and sequencing primers used. A listing of selected multiround PCR amplification and sequencing primers for 13 of the most significant gene loci described here are given in the online supplemental material, and details of the numerous additional PCR primers can be obtained from G. S. Hayward.

Nucleotide sequence accession numbers. A total of 54 new or expanded DNA sequence data files generated here for targeted genomic loci from EEHV3, EEHV4, EEHV5A, EEHV5B, and EEHV6 have been deposited at NCBI GenBank under accession numbers JF692762, JN983092 to JN983126, JX011013, JX011029, JX011085, KC854716 to KC854722, KC854726, KC854743, KC854744, and KC854748 to KC854753. Full details of each locus and their associated accession numbers are listed in Table S1 in the supplemental material. For completeness, data from five additional unchanged files (indicated by an asterisk) reported previously are also included.

RESULTS

Overall sequencing strategy. In the preceding accompanying paper (10), we reported an analysis of the results of extensive Sanger DNA sequencing that generated a total of 378 kb of EEHV genomic DNA sequence derived directly from pathological necropsy tissue samples from eight different elephants that suffered from fatal acute EEHV-associated hemorrhagic disease. These involved three strains of the EEHV1A subgroup and four of the EEHV1B subgroup from Asian elephants and one strain of EEHV2 from an African elephant calf. The data encompassed a total of 57 identifiable open reading frames (ORFs), including 32 that encoded versions of true core proteins common to all herpesviruses, as well as 14 shared with some other herpesviruses and 11 novel genes not found in any other known herpesviruses. On the basis of phylogenetic and genome organization analyses, we proposed that EEHV1 and EEHV2 represented distinct species within the *Proboscivirus* genus and that they might be better considered the prototypes of a newly proposed *Deltaherpesvirinae* subfamily of mammalian herpesviruses rather than their previously assigned status as outliers of the *Betaherpesvirinae* subfamily (10). The recently completed next-generation sequencing-based genome analysis of very high quality necropsy DNA samples from three more strains of EEHV1 (13, 14) has also revealed another 40 or so novel genes in both EEHV1A and EEHV1B, but these genes lie outside the areas covered in the current studies. To further evaluate the four other known major types of EEHV genomes, we carried out selective PCR DNA sequencing to generate between 4 and 32 kb each from six more viremic disease cases (Table 1) representing the prototypes of EEHV3, EEHV4, EEHV5, and EEHV6.

EEHV3 and EEHV4 form a distinctive GC-rich branch of the proboscivirus. The only previous DNA sequence data available for the single known examples each of EEHV3 and EEHV4 (3, 4) was for small highly conserved U60(TERex3, 316-bp) and U38(POL, 480-bp) PCR fragments. While the closest matches to these in the GenBank database were the prototype EEHV1 (Kumari) and EEHV2(Kijana) versions, they still differed from them both at the nucleotide level by 17 to 22% in terminase (TER) and 32 to 35% in DNA polymerase (POL). Furthermore, the closest identity values in BLAST searches for this segment of EEHV3 TER among other herpesviruses are to tupaia herpesvirus (HVtupaia) (70% diverged at the DNA level) and to Old World monkey cytomegaloviruses (CMVs) (55 to 57% diverged at the protein level), whereas for the EEHV3 POL segment, the closest matches are to primate lymphocryptoviruses (LCVs) including Epstein-Barr virus (EBV) (73 to 77% diverged at the DNA level) and to the rhadinovirus retrofibromatosis virus 1 (RFMV1) (56% diverged at the protein level).

Additional DNA sequence data were generated using a large series of experimental redundant two-round and three-round nested consensus PCR primers for other targeted gene loci. Despite designing these primers based on anticipated residual homology between the known EEHVs and other herpesviruses, this process proved to be extremely difficult and overall yielded a total of just 4 to 5 kb each over five successful PCR loci, with the largest locus being 1,250 bp from the DNA polymerase. Few of the redundant consensus primers that worked here for one of these two new viruses were successful directly for both, but in each case, the sequence data obtained from the first virus was used to then generate additional appropriate internal primers for PCR amplifying and sequencing the equivalent regions from the other virus. The combined sequence data across all five PCR loci identified part or all of eight predicted EEHV3 and EEHV4 ORFs as shown in abbreviated summary form in Table 2 and listed in greater detail in Table S1 in the supplemental material. Their map locations relative to the complete genome of EEHV1A(Kimba) are shown in the schematic diagram in Fig. 1. The data obtained proved sufficient for relatively robust evaluation of the genetic differences and phylogenetic relationships at the nucleotide and protein levels for segments of three of the most conserved genes POL, portal (POR), and helicase (HEL), as well as for the far more diverged U71(gM) and origin binding protein (OBP) loci. However, the TER locus showed too little amino acid divergence (just 3 to 5%) between EEHV species in this region to allow a meaningful phylogenetic comparison (although that does not apply for TER at the DNA level). A diagrammatic illustration of the huge divergence of both EEHV3 and EEHV4 from the other known EEHV types, as well as their close relationship to each other, is presented in Fig. 2, where a 1,077-bp DNA segment of U38(POL) that is common to all eight prototype genome types is evaluated in a distance-based Bayesian phylogenetic tree. Within this subregion, the EEHV3 and EEHV4 POL genes differ from one another by just 4.0%.

The major reason that it proved so difficult to develop consensus primers for EEHV3 and EEHV4 based on the EEHV1 and EEHV2 DNA sequence data was evidently because of a huge difference in G+C content between these two branches of EEHVs. For these five segments of the EEHV3 and EEHV4 genomes at least, the G+C content of EEHV3 and EEHV4 ranges between 56 and 67% compared with overall values between 41 and 43% for all six AT-rich branch genomes. The effect is concentrated within the third "wobble" position of most coding triplets, and this feature applies to all seven EEHV3 and EEHV4 proteins analyzed as listed in Table 3**.** Several of even the most conserved genes (e.g., TER, POR, and HEL) have acquired nearly the maximum possible G+C content bias of between 96 and 99% G+C content in the third wobble position. This compares to values of between 41 and 50% in the wobble codon position for the same protein regions in their EEHV1A orthologues. The tendency for some herpesvirus genomes to display an extraordinarily high GC bias, and others to have a high AT bias, is found among several genera or lineages from all three mammalian subfamilies. The most extreme example of unusually high G+C content is pseudorabies virus (PRV) at 72%, whereas unusually low G+C content down to between 42 and 46% is found in varicella-zoster virus (VZV), herpesvirus saimiri (HVsaimiri) (HVS) (unique segment), human herpesvirus 6 (HHV6), and now also EEHV1.

TABLE 2 Summary of PCR-sequenced EEHV3, EEHV4, EEHV5A, EEHV5B, and EEHV6 gene coding regions

Gene/ORF no./ID[a]	HCMV ORF	HSV ORF	Orientation[b]	Protein name	Status[c]	Presence/absence of the indicated ORF[d]					
						EEHV3 NAP27	EEHV4 NAP22	EEHV5A NAP28	EEHV5A NAP50	EEHV5B NAP58	EEHV6 NAP35
U39	UL55	UL29	F	gB	Core		+	+	+	+	+
U38	UL54	UL30	F	POL	Core	+	+	+	+	+	+
U33	UL49	Nil	F	Cys-rich	β/γ			+	+	+	+
U28	UL45	UL39	F	RRA	Core			+	+	+	+
U27.5/ORF-H	Nil	UL40	F	RRB	α/β2			+	+	+	+
U27/ORF-I	UL44	UL42	F	PPF	Core				+	+	+
U45.7/ORF-J	Nil	Nil	F		Novel						+
U46	UL73	UL49A	F	gN	Core						+
U47/ORF-D	UL74	Nil	R	gO	β						+
U48	UL75	UL22	R	gH	Core			+	+	+	+
U48.5/ORF-E	Nil	UL23	R	TK	α/γ			+	+	+	+
U49	UL76	UL24	F		Core			+	+	+	+
U50	UL77	UL25	F	PAC2	Core				+	+	+
U51	UL78	Nil	F	vGPCR1	β			+	+	+	+
U57	UL86	UL19	R	MCP	Core			+	+	+	+
U60ex3	UL89	UL15ex2	R	TERex3	Core	+	+		+	+	+
U62	UL91	Nil	F		β/γ				+	+	+
U63	UL92	Nil	F		β/γ						+
U66ex2	Nil	Nil	R	TERex2	Core						+
U66ex1	UL89ex1	UL15ex1	R	TERex1	Core						+
U70	UL98	UL12	F	EXO	Core	+	+	+	+	+	+
U71	UL99	UL11	F	myrTeg.	Core	+	+	+	+	+	+
U72	UL100	UL10	R	gM	Core	+	+	+	+	+	+
U73/ORF-G	Nil	UL09	F	OBP	α/β	+	+	+	+	+	+
U76	UL104	UL06	R	POR	Core	+	+	+	+	+	+
U77	UL105	UL05	F	HEL	Core	+	+	+	+	+	+
U77.5/ORF-M	Nil	Nil	F	Nuclear	Novel				+	+	+
U80.5/ORF-N	Nil	Nil	R	vCXCL1	Novel				+	+	+
U81	UL114	UL02	R	UDG	Core				+	+	+
U82	UL115	UL01	R	gL	Core				+	+	+
U82.5/ORF-Oex3	Nil	Nil	R	S/TGlyP	Novel				+	+	+
U82.5/ORF-Oex2	Nil	Nil	R	S/TGlyP	Novel				+	+	+
U82.5/ORF-Oex1	Nil	Nil	R	S/TGlyP	Novel				+	+	+
U83.5/ORF-Pex2	Nil	Nil	R	S/TGlyP	Novel					+	
U83.5/ORF-Pex1	Nil	Nil	R	S/TGlyP	Novel					+	
U84.5/ORF-Q	Nil	Nil	R	S/TGlyP	Novel					A	
U85.5/ORF-Kex3	Nil	Nil	R	SplGlyP	Novel				+	+	+
U85.5/ORF-Kex2	Nil	Nil	R	SplGlyP	Novel				+	+	+
U85.5/ORF-Kex1	Nil	Nil	R	SplGlyP	Novel				+	+	+
U86.5/ORF-L	Nil	Nil	R	IE-like	Novel				+	+	+

[a] ID, identification.
[b] F, forward; R, reverse.
[c] Novel, not found in any other herpesviruses; β, betaherpesvirus subfamily only; Core, common to all herpesvirus subfamilies; β/γ, betaherpesvirus and gammaherpesvirus subfamilies only; α/β, alphaherpesvirus and betaherpesvirus subfamilies only; α/β2, alphaherpesvirus subfamily and roseoloviruses only.
[d] +, partial or intact ORF present (see Table S1 in supplemental material for detailed coordinates, percent divergence, and GenBank accession numbers); A, gene absent.

Comparisons of conserved ORFs within the EEHV3 and EEHV4 genomes. A summary of the DNA and protein level identity relationships (expressed as percentage differences) among the seven largest individual ORF segments in common between EEHV3 and EEHV4 compared to EEHV1A (10) is provided in Table 3. The ORFs are listed in projected order from left to right across the genome and with provisional equivalent map coordinates given based on where they match to the recent complete draft 177,316-bp DNA sequence now available for EEHV1A (Kimba) as reported by Ling et al. (GenBank accession no. KC618527) (13). The observed nucleotide level differences for EEHV3(NAP27) ORFs compared to EEHV1A(Kumari) for the four PCR segments encompassing POL, OBP, POR, and HEL range from 30 up to 48% (but with TER being just 22%). These values are considerably larger than the 20 to 27% for matching segments of the same four ORFs of EEHV1A(Kumari) compared to EEHV2(Kijana). Furthermore, EEHV3 and EEHV4 themselves differ at the DNA level by 7.0, 17, 3.3, 3.5, and 3.8% within their matching sequenced segments of POL, OBP, POR, HEL, and TER, respectively. Therefore, while the evolutionary divergence of EEHV3 from EEHV4 in the OBP and POL loci would certainly be sufficient to argue for separate species status, the differences are nevertheless quite low for the more highly conserved POR, HEL, and TER loci.

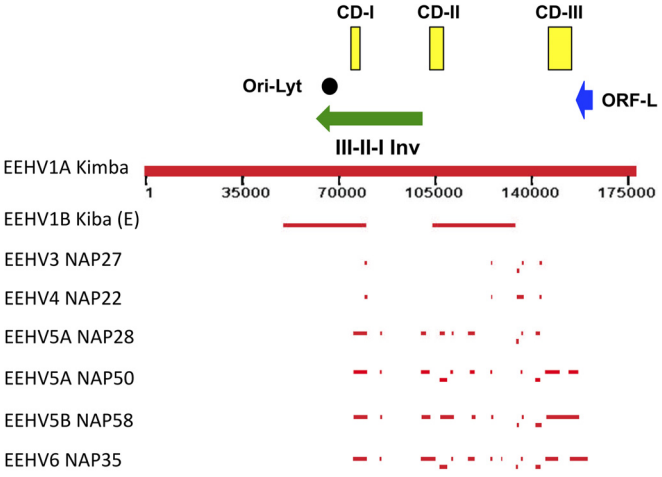


FIG 1 Schematic map of position coordinates for all sequenced loci in EEHV3, EEHV4, EEHV5A, EEHV5B, and EEHV6 compared to the complete EEHV1A(Kimba) genome. The diagram is drawn to scale with bars representing all of the DNA sequence blocks generated here aligned relative to the complete 177,316-kb genome of EEHV1A(Kimba) (13). The data reported by Ehlers et al. (11) for EEHV1B(Kiba) is also given for comparison in the top line. The locations of the predicted Ori-Lyt dyad symmetry locus (black circle), the large 40-kb inverted (Inv) core domain III-II-I segment (green arrow), the putative immediate early-like ORF-L transactivator protein coding region (blue arrow), and the three major hypervariable domains CD-I, CD-II, and CD-III (yellow boxes) as described by Richman et al. (10) are indicated.

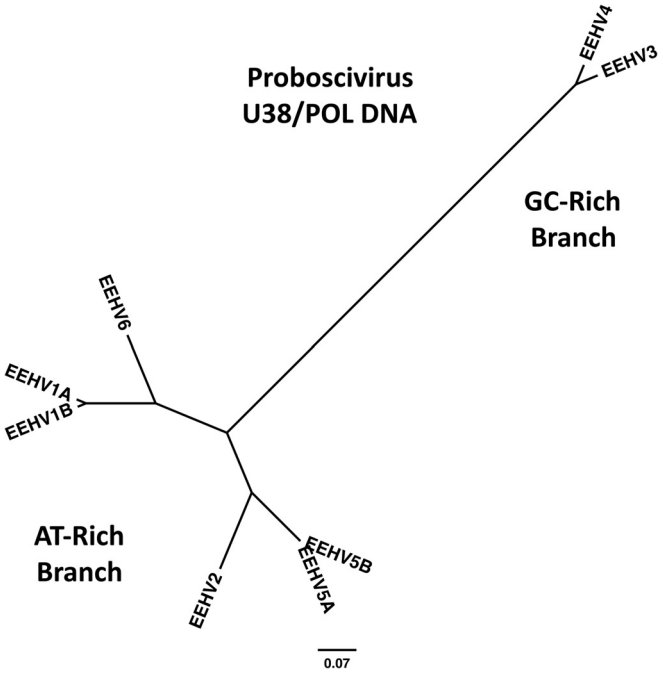


FIG 2 Radial phylogenetic tree showing evolutionary relationships between EEHV1, EEHV2, EEHV3, EEHV4, EEHV5, and EEHV6 in the highly conserved U38(POL) DNA gene. The diagram shows a distance-based Bayesian nearest-neighbor evolutionary tree dendrogram produced in MEGA5 illustrating the branching patterns of the GC-rich (EEHV3 plus EEHV4) compared to the AT-rich EEHV groups and of the two distinct subgroups (EEHV1 plus EEHV6 compared to EEHV2 and EEHV5) within the AT-rich branch. After several small gaps and nonaligned nucleotides were omitted, the final data set for the segment of U38(POL) gene DNA used that is common to all eight EEHV types was 1,077 bp. The number of nucleotide substitutions per site is shown by the bar.

Comparison of EEHV3 and EEHV4 across the highly divergent U71-gM PCR locus. In addition, a fifth PCR segment obtained encodes all of the small but highly diverged U71(MyrTeg) protein inserted between the convergent C termini of the U70(EXO) and U72(gM) genes. This region was accessible because an overlap between the coding regions at the C terminus of the U70 gene with the N terminus of U71 in two different reading frames creates a very highly conserved 60-bp nucleotide block that proved suitable as a template for multiple adjacent conserved PCR primers. Nevertheless, the U71 proteins from EEHV3 and EEHV4 differ from the EEHV1(Kumari) versions by 63% and 66% at the amino acid level, respectively. Furthermore, the intact U71 and partial U72 ORFs obtained for EEHV3 and EEHV4 differ from each other here by 26% and 14% at the DNA level and by 29% and 13% at the amino acid level. The lengths of the intergenic noncoding segment between the tail-to-tail oriented C termini of U71 and U72 are 165 bp for EEHV3 and 194 bp for EEHV4. Both intergenic regions contain several homonucleotide tracts such as three TTT TTTTT (T8) tracts and one AAAAAAAA (A8) tract in EEHV3 and 1× T8 tract, 1× T7 tract, and 1× A6 tract in EEHV4, but there is little nucleotide identity between them otherwise.

Catenation of the sequences from all five analyzed PCR loci for EEHV3 compared to matching segments of EEHV4 shows 462-bp differences out of a total of 3,708 bp in common between them (12.5%). However, when the hypervariable U71-gM locus is omitted, the DNA difference values for the sum of the four remaining core gene loci are just 250 out of 2,913 bp (8%). Finally, for EEHV4 only, we were also able to use a long-range PCR approach to join across the gap between the U72(EXO) and U73(OBP) loci and increase the total to 5,743 bp. The data revealed an additional inserted 720-bp region here compared to EEHV1 consisting of a novel 480-bp intergenic domain containing 13 more 5- to 10-bp-long A or T homonucleotide runs, as well as a GC-rich extension adding an 82-amino-acid (aa) nonhomologous region at the N terminus of the OBP protein. This result accounts for the large increase shown in Table S1 in the supplemental material in overall nucleotide differences from EEHV1 for all EEHV4 loci (48%) compared to EEHV3 (37%).

Evaluation of two more AT-rich branch probosciviruses EEHV5 and EEHV6. The initial DNA sequence data available for the first strains identified of EEHV5 and EEHV6 were just from the small POL codehops regions (4), which both showed 21% divergence from their closest orthologues (EEHV2 or EEHV1) and 27% from each other. Later addition of the U71-gM locus suggested the possibility of two distinct subgroups of EEHV5 (17). To obtain more robust phylogenetic data from several clinical samples, extensive initial trial and error testing was again carried out with large numbers of partially redundant two- and three-round nested PCR primers designed based on conserved sequence features between EEHV1 and EEHV2. Selected subsets of successful primers were then used in judicious combinations with second-generation EEHV5- or EEHV6-specific primers. The process of identifying successful primers was more efficient this time, because although there was less viral DNA present, they were less divergent from the reference EEHV1 and EEHV2 sequences. Eventually, as summarized in Table 1, we successfully accumulated a total of 31.8 kb from 13 PCR loci for EEHV6(NAP35), as

TABLE 3 DNA and protein divergence in ORFs across EEHV3, EEHV4, and EEHV1A

Gene locus	EEHV1A coordinates[a]	Protein size (no. of aa)[b]	Nucleotide level divergence (%)[c]			Amino acid level divergence (%)[c]			Overall % G+C content[d]			Wobble % G+C content[d]		
			3/4	3/1A	4/1A	3/4	3/1A	4/1A	3	4	1A	3	4	1A
U38(POL)	(77782–78912)	(412)	7.0	34	34	7.5	26	25	67	68	42	89	89	43
U60(TERex3)	(123721–124037)	(105)	3.8	22	23	1	5	5	57	56	42	91	86	41
U71(MyrTeg)	132954–133241	96	26	50	49	29	63	66	61	58	46	76	61	41
U72(gM)	(133320–133608)	(95)	14	45		13	52		55		40	80		42
U72(gM)	(133316–134404)	372			34			35		55	40		89	49
U73(OBP)	(134645–135283)	(237)	17	48		22	39		63		41	91		46
U73(OBP)	(134404–135284)	(438)			54			54		64	44		88	49
U76(POR)	(140967–141295)	(104)	3.3	34	30	1	21	21	64	63	42	98	99	47
U77(HEL)	(141246–141878)	(189)	3.5	30	30	2.5	23	22	67	67	41	96	97	50

[a] The EEHV1A coordinates are based on Ling et al. (13). Entries shown in parentheses indicate that an incomplete ORF or smaller region was used.
[b] aa, amino acids.
[c] Divergence of EEHV3, EEHV4, and EEHV1A is indicated as follows: EEHV3 and EEHV4 (3/4), EEHV3 and EEHV1A (3/1A), and EEHV4 andEEHV1A (4/1A).
[d] The total guanine-plus-cytosine nucleotide percentages of EEHV3, EEHV4, and EEHV1A (indicated by 3, 4, and 1A in the table) compared to those at just the third nucleotide position in all codons from the predicted proteins.

well as 14.6, 26.1, and 29.3 kb from 9 to 12 loci each for EEHV5A(NAP28), EEHV5A(NAP50), and EEHV5B(NAP58). For both virus types, this should represent close to 15% of their total genomes, considering the observed nearly180-kb size for EEHV1 (13, 14). A schematic summary diagram showing the map locations of all of these sequenced loci relative to the intact genome of EEHV1A(Kimba) is presented in Fig. 1. Interpretations of all the partial or complete predicted ORFs identified are summarized in Table 2, whereas Table S1 in the supplemental material lists in greater detail the ORF positions and sizes, including GenBank file accession numbers and matching coordinates and levels of divergence from EEHV1(Kimba) for all 54 DNA loci evaluated here.

The results indicated that the genomes of EEHV5 and EEHV6 are again much more closely related to members of the *Proboscivirus* genus than to any other known herpesviruses at all loci. In addition, they both belong to the AT-rich branch with an overall base composition of close to 43% G+C content and with an average of just below 50% G+C content at the third wobble codon position across all loci tested. Furthermore, they have both diverged at a relatively uniform pattern all the way across the conserved core region of their genomes, with each EEHV6 locus resembling EEHV1 more than EEHV2, whereas all EEHV5 DNA loci more closely resembled EEHV2 than EEHV1. A simple initial illustrative radial phylogenetic tree for 1,080 bp of POL DNA illustrates the relative evolutionary positions of EEHV5A, EEHV5B, and EEHV6 compared to the other five EEHV types within both the AT-rich and GC-rich branches (Fig. 2).

Divergence of EEHV6 proteins compared to other EEHVs. Overall, the EEHV6 genome data obtained encompass 31 ORFs that all have orthologues in EEHV1 and EEHV2 and show the same gene organization relative to their immediate neighbors. EEHV6 also has the same novel polymerase processivity factor (PPF) to glycoprotein N (gN) core gene block III-II-I inversion junction, and both encode typical thymidine kinase (TK), ribonucleotide kinase B subunit (RRB), viral G-protein-coupled receptor 1 (vGPCR1), OBP, ORF-J, ORF-M, ORF-N(vCXCL1) (vCXCL1 stands for viral CXC chemokine ligand 1), ORF-O, ORF-K, and ORF-L(immediate early [IE]-like) proteins that are unique to the *Proboscivirus* genus.

Comparisons of the DNA and protein level identity relationships (expressed as percentage differences) versus previous data obtained for EEHV1A and EEHV1B (10) is provided in Table 4 for the 29 largest EEHV6 ORFs evaluated, which are listed in projected order across the genome with equivalent positional coordinates given based on EEHV1A(Kimba) (GenBank accession no. KC618527) (13). All ORFs in the chimeric hypervariable regions of EEHV1A and EEHV1B that show between 8 and 35% DNA level divergence from each other are shown in bold type. The 13 most conserved EEHV6 genes differ at the DNA level from their closest orthologues in EEHV1A by between 11 and 17%, with 8 more genes giving values between 18 and 23%, whereas the divergence reaches 25 to 31% for myristylated tegument protein (MyrTeg), ORF-K, and ORF-N(vCXCL1). Similarly, within the most conserved group, nine ORFs (not counting TER) show protein level differences ranging from just 4.2 to 9.2% that are well below their DNA level differences, whereas for 21 others, the protein divergence values of 11 to 26% fairly closely match the DNA values. ORF-N(vCXCL1) shows the highest protein divergence at 29%. Finally, the EEHV6 data includes a large segment of the novel U86.5(ORF-L) gene encoding a predicted IE-like nuclear transactivator protein that differs from those of both EEHV1A and EEHV1B by 18% at the protein level.

EEHV6 is more closely related to EEHV1A than to EEHV1B within the chimeric gene blocks. An unusual and unexpected pattern emerged when comparing the EEHV6 genomes with those of EEHV1B across the three major chimeric domains referred to as chimeric domain I (CD-I), CD-II, and CD-III (10) (Table 4). Evidently, whereas EEHV1A and EEHV6 are fairly uniformly diverged all the way across the sampled core 85-kb segment of their genomes, including across the equivalents of the chimeric portions, this is not the case for EEHV1B. In particular, EEHV1B DNA differs substantially more from EEHV1A than EEHV6 differs from EEHV1A in both the CD-II (32% versus 16%) and CD-III (37% versus 17%) blocks, values that are nearly as high as the 35 to 42% found between EEHV1 and EEHV2 here. Furthermore, the amino acid differences for the six least conserved EEHV6 proteins (ORF-J, gO, gH, uracil DNA glycosylase [UDG], gL, and ORF-O) from the EEHV1B versions reach between 30 and 41%,

TABLE 4 DNA and protein divergence in ORFs across the EEHV1A, EEHV1B, and EEHV6 genomes[a]

Gene locus	EEHV1A(Kimba) coordinates[b]	Protein size[b]	Nucleotide level divergence (%)[c]			Amino acid level divergence (%)[c]			Chimeric domain
			1A-1B	1A-6	1B-6	1A-1B	1A-6	1B-6	
U39(gB)	**73959–76511**	**836**	**21**	**16**	**20**	**14**	**11**	**13**	**CD-I**
U38/(POL)	(76544–79043)	(832)	5	17	16	4	12	12	CD-I
U33	(83628–84428)	(266)	1.3	22	21	0.3	19	18	
U28(RRA)	99358–99760	801	1.1	14	15	0	8	8	
U27.5(RRB)	99804–100709	302	1.4	14	14	0.3	17	17	
U27(PPF)	100960–102186	408	1.8	16	16	0.5	6	6	
U45.7(ORF-J)	**102168–102775**	**168**	**26**	**16**	**25**	**32**	**14**	**33**	**CD-II**
U46(gN)	**102759–103713**	**96**	**29**	**16**	**28**	**20**	**6.2**	**18**	**CD-II**
U47(gO)	**103075–103713**	**212**	**35**	**16**	**35**	**38**	**11**	**37**	**CD-II**
U48(gH)	**(105363–105903)**	**(179)**	**31**	**17**	**28**	**33**	**21**	**37**	**CD-II**
U48.5(TK)	**105835–106908**	**356**	**13**	**16**	**18**	**16**	**21**	**20**	**CD-II**
U49	106910–107608	232	3.6	15	15	0.4	12	13	
U50(PAC2)	(107427–108083)	(220)	4.0	23	23	3.1	20	21	
U51(vGPCR1)	(109398–110239)	(286)	4.2	22	23	3.3	14	17	
U57(MCP)	(115577–117866)	(748)	2.7	17	17	1.2	7	7	
U60(TERex3)	123601–124175	(194)	2.9	12	13	0[d]	0[d]	0[d]	
U62	124231–124477	88	0.8	13	13	0	9	9	
U71(MyrTeg)	**132954–133241**	**96**	**8**	**25**	**25**	**9**	**23**	**26**	
U72(gM)	(133320–133614)	(97)	4.4	13	14	3.3	9.2	7.2	
U73(OBP)	(134615–135415)	(277)	3.0	14	14	1.9	6.0	6.7	
U76(POR)	(139998–141295)	(433)	1.9	13	13	0	4.2	4.2	
U77(HEL)	(141294–141864)	(190)	2.8	12	13	0.5	7.0	7.5	
U77.5(ORF-M)	**143988–145502**	**503**	**16**	**16**	**19**	**13**	**13**	**16**	**CD-III**
U80.5(ORF-N)	**145642–145959**	**106**	**Del**	**31**	**Del**	**Del**	**29**	**Del**	**CD-III**
U81(UDG)	**146027–146980**	**317**	**29**	**19**	**30**	**29**	**19**	**30**	**CD-III**
U82(gL)	**146946–147860**	**304**	**31**	**18**	**30**	**33**	**22**	**41**	**CD-III**
U82.5(ORF-O)	**(147700–147910)**	**(78)**	**33**	**16**	**31**	**34**	**18**	**37**	**CD-III**
U85.5(ORF-K)	(152200–153350)	(380)	1.7	26	26	0.8	25	25	
U86.5(ORF-L)	(155324–158615)	(1084)	2.4	11	11	0.6	18	18	

[a] The rows in the table shown in bold type indicate ORFs that are highly variable between EEHV1A and EEHV1B.
[b] The EEHV1A coordinates are based on Ling et al. (13). The values shown in parentheses indicate that an incomplete ORF or shorter region was used.
[c] The divergence between EEHV1A and EEHV1B (1A-1B), EEHV1A and EEHV6 (1A-6), and EEHV1B and EEHV6 (1B-6) is shown. Del, ORF-N is absent (deleted) in all EEHV1B strains.
[d] All three TER proteins are identical or only 1 aa different over this 194-aa segment.

much higher than the differences from their EEHV1A counterparts.

Even within the first chimeric block of EEHV1B(CD-I) of 3.0 kb in size (Kimba coordinates 74,016 to 77,012) encompassing parts of both U39(gB) and U38(POL), the nucleotide difference values for EEHV1A-EEHV1B (EEHV1A-1B), EEHV6-EEHV1A (EEHV6-1A), and EEHV6-EEHV1B (EEHV6-1B) are all nearly the same at 21%, 17%, and 21%, respectively, whereas all three of them differ from EEHV2 here by between 26 to 29%. This is illustrated pictorially in the SimPlot diagrams shown in Fig. 3a to c, where there is a dramatic crossover between the similarity lines for EEHV1A versus EEHV1B compared to those for EEHV1A versus EEHV6 at position 3,000 (indicated by an arrow) representing the right-hand side boundary of CD-I for both Kiba and Haji (Fig. 3b and c). In contrast, the lines for EEHV1A versus EEHV2 and for EEHV1A versus EEHV6 are relatively uniform all the way across the 5-kb segment shown (Fig. 3a).

SimPlot comparisons of the patterns of similarity and divergence across the 3.7-kb CD-II block (Kimba coordinates 102,509 to 106,186) for EEHV6 versus EEHV1A and EEHV1B are shown in Fig. 4. Separate segments of 2.6 and 2.7 kb encompassing both sides of the region are shown, although data for the central 1.6-kb segment of the EEHV6 U48(gH) gene between these two blocks is not yet available. However, the five adjacent genes, ORF-J, gN, gO, gH (C terminus), and TK, within the CD-II equivalent region of EEHV6 all proved to be significantly more closely related to EEHV1A than to EEHV1B. Specifically, when the left-hand side (Fig. 4a) and right-hand side (Fig. 4b) boundary segments of CD-II are displayed for EEHV1A versus EEHV6 in comparison to EEHV1A versus EEHV1B, both revealed a sharp crossover transition, where the divergence between EEHV1B and EEHV1A changes to become much greater than that between EEHV1A and EEHV6. For example, comparing the 1.3 kb to the right of and inside the left-hand side chimeric junction at position 1,330 within the ORF-J—gN–gO gene block (indicated by an arrow in Fig. 4a), the overall nucleotide difference values for EEHV1A-1B, EEHV6-1A, and EEHV6-1B are 33%-16%-31%. That compares with 2.5%-16%-17% for the next adjacent 1.3 kb on the left just outside the boundary. Similarly, over the U48(gH)-U48.5(TK) block just inside the right-hand side chimeric boundary (indicated by an arrow at position 850 in Fig. 4b), the differences for EEHV1A-1B, EEHV6-1A and EEHV6-1B are 29%-16%-24%, whereas over the next 1.8 kb to the right in the adjacent outside flanking U49-U50 gene region, the values are 3.5%-15%-15%.

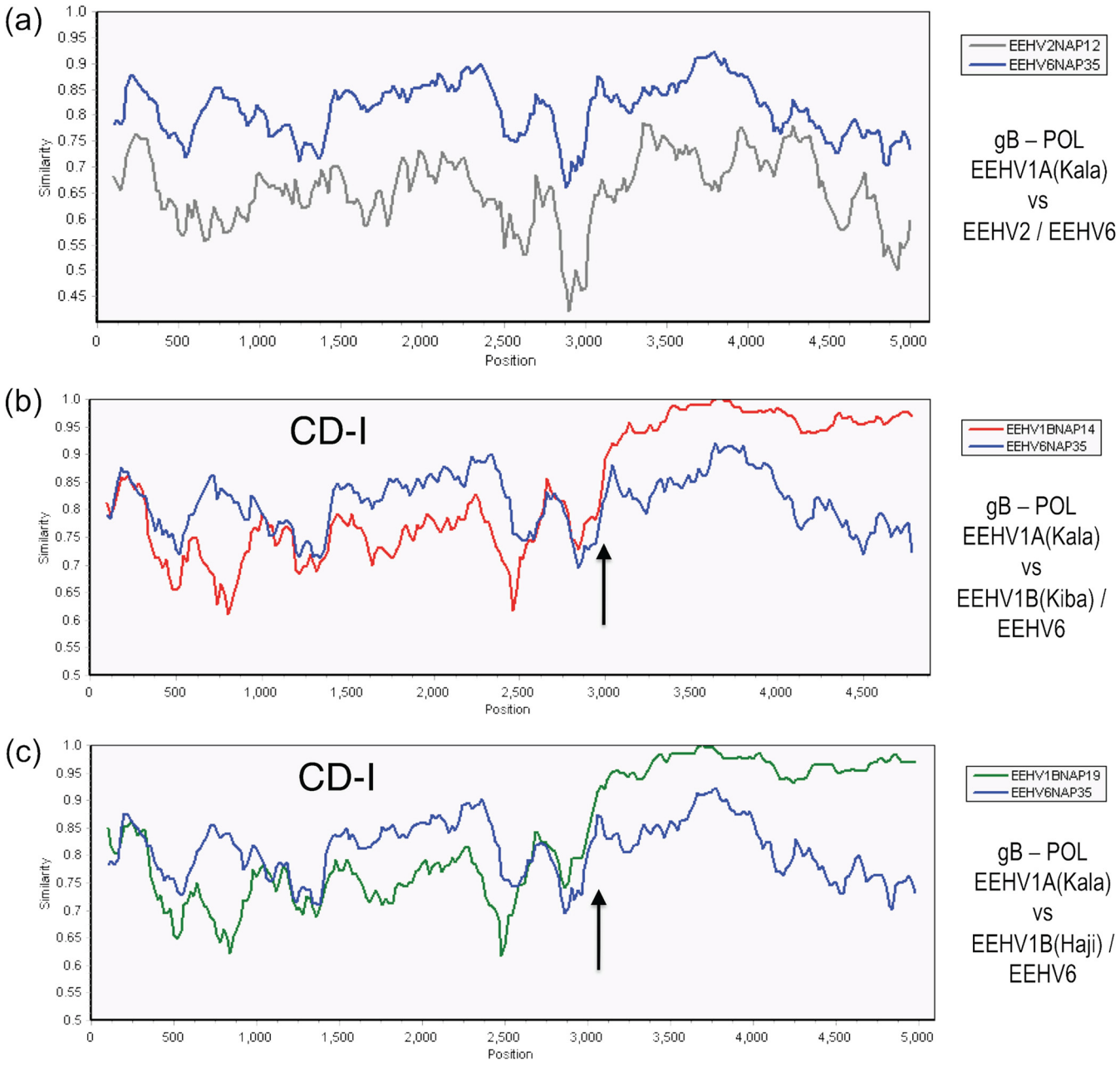


FIG 3 Evaluation of EEHV1A-1B chimeric domain CD-I patterns and boundaries relative to EEHV6. The diagrams show SimPlot comparisons of the nucleotide identity patterns between EEHV6, EEHV1A, EEHV1B, and EEHV2 across the 3.0-kb EEHV1B chimeric domain CD-I. (a) CD-I. The 5,000-bp U39(gB)-U38(POL) segment from EEHV1A(Kala, NAP18) map coordinates 73,959 to 79,043 compared to EEHV6(NAP35) (blue) and to EEHV2(Kijana, NAP12) (gray) is shown. (b) CD-I. The 4,800-bp U39(gB)-U38(POL) segment from EEHV1A(Kala, NAP18) map coordinates 73,987 to 78,860 compared to EEHV1B(Kiba, NAP14) (red) and to EEHV6(NAP35) (blue) is shown. (c) CD-I. The 5,000-bp U39(gB)-U38(POL) segment from EEHV1A(Kala, NAP18) map coordinates 73,959 to 79,043 compared to EEHV1B(Haji, NAP19) (green) and to EEHV6(NAP35) (blue) is shown. Arrows mark the positions of the chimeric domain boundary transitions, and the relevant DNA accession numbers are included in Table S1 in the supplemental material.

Finally, within the third major chimeric locus of EEHV1B (CD-III, 8.3 kb, Kimba coordinates 143,644 to 152,026), the DNA difference values for EEHV1A-1B, EEHV6-1A, and EEHV6-1B are again plotted across a contiguous 4.3-kb region from U77(HEL) to ORF-O that encompasses the left-hand side boundary (Fig. 4c). After typical values of 2.8%-12%-13% for U77 (HEL), the analysis then reveals nucleotide level differences of 16%-16%-19% for ORF-M and even higher differences at 29%-19%-30%, 31%-18%-30%, and 33%-16%-31% for UDG, gL, and ORF-O, respectively (Table 4). The bulk of ORF-M is not considered part of the chimeric domain, but instead it is interpreted to begin after the sharp transitional boundary (indicated by an arrow) at position 1,600 in Fig. 4c. Although the rest of ORF-O, as well as ORF-P and ORF-Q and the corresponding right-hand side boundary, are not available yet for EEHV6, the values return to the expected more typical pattern of 1.7%-26%-26% and 2.4%-11%-

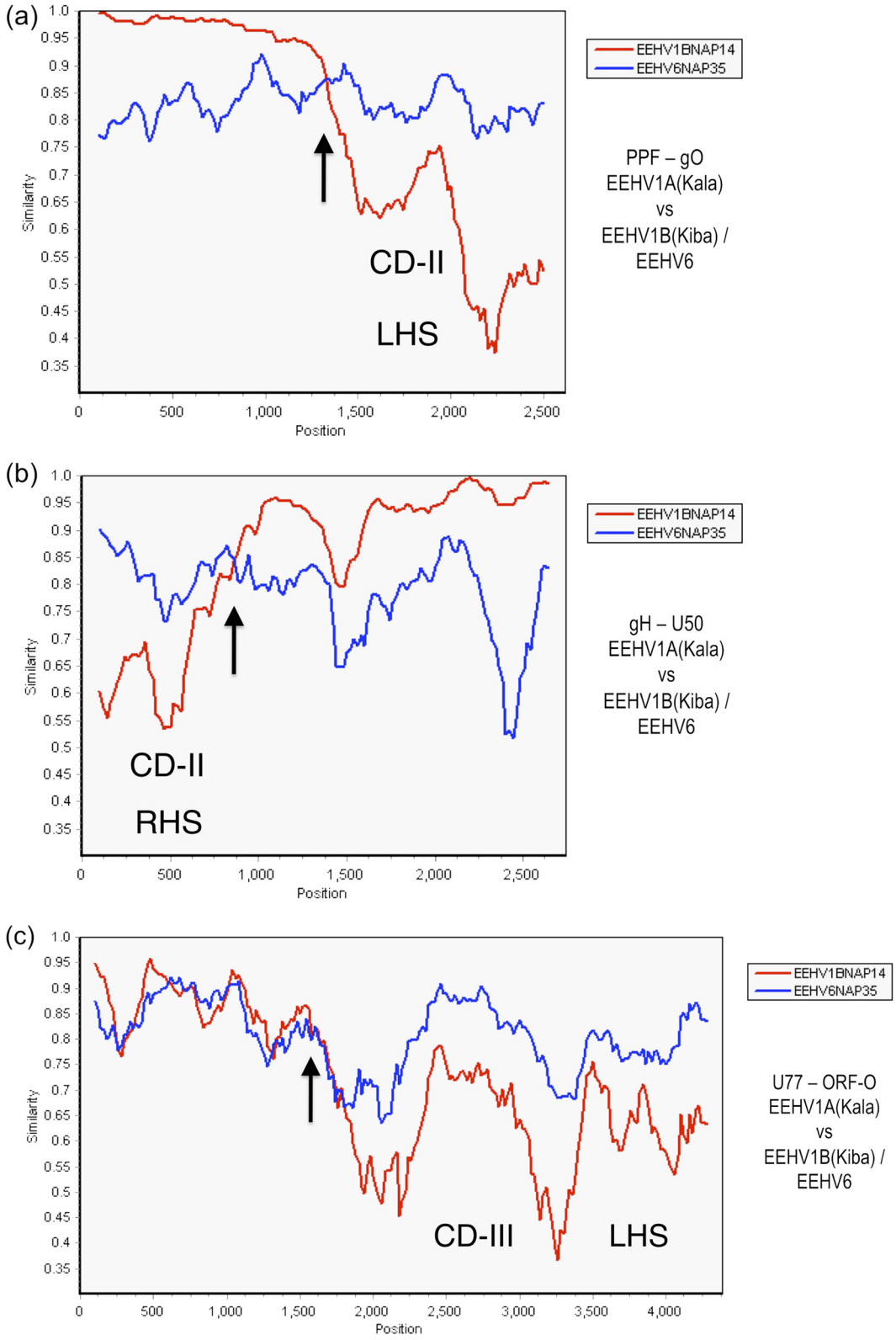


FIG 4 Evaluation of EEHV1A-1B chimeric domain CD-II and CD-III patterns and boundaries relative to EEHV6. The diagrams show SimPlot comparisons of the nucleotide similarity patterns between EEHV6, EEHV1A, and EEHV1B across both the left-hand side (LHS) and right-hand side (RHS) boundaries of the 3.7-kb CD-II region and across the LHS boundary only of the 8.3-kb CD-III region of EEHV1B. (a) CD-II LHS. The left-hand side of the 2.6-kb U27(PPF)-U47(gO) segment across map coordinates 101,128 to 103,703 for EEHV1A(Kala, NAP18) compared to EEHV1B(Kiba, NAP14) (red) and to EEHV6(NAP35) (blue). (b) CD-II RHS. The right-hand side of the 2.7-kb U48(gH)-U50(PAC2) segment across map coordinates 105,364 to 108,084 for EEHV1A(Kala, NAP18) compared to EEHV1B(Kiba, NAP14) (red) and to EEHV6(NAP35) (blue) is shown. (c) CD-III LHS. The left-hand side of the 4.3-kb U77-U82.5(ORF-O) segment across map coordinates 143,644 to 147,936 for EEHV1A(Kala, NAP18) compared to EEHV1B(Kiba, NAP14) (red) and to EEHV6(NAP35) (blue) is shown. (Data for the right-hand segment of CD-III are not yet available for EEHV6.) Arrows mark the positions of the chimeric domain boundary transitions, and the relevant DNA accession numbers are included in Table S1 in the supplemental material.

TABLE 5 DNA and protein divergence in ORFs across EEHV5A, EEHV5B, EEHV2, and EEHV1A[a]

Gene locus	EEHV1A coordinates[b]	Protein size[b]	Nucleotide level divergence (%)[c]				Amino acid level divergence (%)[c]				Chimeric domain
			5A-5B	5A-2	5B-2	5A-1	5A-5B	5A-2	5B-2	5A-1	
U39(gB)	**(74091–76511)**	**(837)**	**10**	**16**	**15**	**27**	**4.2**	**7.6**	**6.3**	**21**	**CD-I**
U38(POL)	(76583–79038)	(818)	1.3	18	17	24	1.4	11	11	21	CD-1
U33	(83648–84183)	(176)	0.8	28	28	37	0	25	25	43	
U28(RRA)	(98637–99760)	(374)	0.4	17	17	23	0	8.6	8.6	16	
U27.5(RRB)	99804–100709	(301)	0.9	16	16	24	0.7	6.3	6.3	15	
U27(PPF)	(100960–101609)	(216)	0.8	19	19	31	0.9	14	14	25	
U48(gH)	(105396–105911)	(172)	0	18	18	38	0	16	16	34	
U48.5(TK)	**105835–106908**	**344**	**14**	**18**	**20**	**30**	**13**	**19**	**19**	**27**	**CD-II**
U49	**106910–107628**	**233**	**6.0**	**11**	**10**	**24**	**2.1**	**6.9**	**4.7**	**24**	**CD-II**
U50(PAC2)	(107428–108178)	(261)	0.2	18	18	32	0	19	19	34	
U51(vGPCR1)	**(109225–110065)**	**(292)**	**4.6**	**22**	**23**	**32**	**2.7**	**17**	**17**	**25**	
U57(MCP)	(116671–117809)	(380)	0.2	19	19	27	0.3	9.2	9.2	19	
U60(TERex3)	(123674–124180)	(169)	1.2	13	13	21	0	1.3	1.3	3.0	
U62	(124231–124434)	(75)	0.5	16	16	29	0	10	10	19	
U71(MyrTeg)	**132954–133241**	**98**	**8.7**	**24**	**23**	**44**	**14**	**27**	**30**	**56**	
U72(gM)	133316–133613	(101)	0.3	16	17	21	0	10	10	12	
U73(OBP)	(134605–134430)	(269)	0	12	12	21	0	4.4	4.4	10	
U76(POR)	(139815–141295)	(493)	0.2	13	13	22	0.4	4.1	4.1	7.8	
U77(HEL)	(141294–141793)	(181)	0.4	14	14	24	0.7	6.0	6.0	14	
U77.5(ORF-M)	143988–145511	484	1.1	16	17	29	0	10	10	17	CD-III
U80.5(ORF-N)	**145642–145959**	**106**	**28**	**23**	**27**	**38**	**32**	**24**	**33**	**47**	**CD-III**
U81(UDG)	**146027–146980**	**324**	**27**	**23**	**28**	**33**	**23**	**21**	**20**	**29**	**CD-III**
U82(gL)	**146946–147776**	**312**	**27**	**17**	**27**	**34**	**26**	**16**	**26**	**44**	**CD-III**
U82.5(ORF-Oex3)	**147694–147958**	**(88)**	**23**	**21**	**25**	**32**	**21**	**20**	**20**	**25**	**CD-III**
U82.5(ORF-O)	147700–148985	487			36	54[d]			34	53[d]	
U83.5(ORF-P)	146932–150638	526			33	48[d]			33	52[d]	
U85.5(ORF-K)	152042–154436	(678)	1.3	23	22	41	0.4	29	27	48	
U86.5(ORF-L)	155053–155356	(97)	1.4	10	12	30	1.0	7.0	8.4	20	

[a] The prototype genomes used were EEHV5A(NAP50), EEHV5B(NAP58), EEHV2(NAP12), and EEHV1A(NAP18). The rows in the table shown in bold type indicate ORFs with high-level variability (hypervariable genes) between EEHV5A and EEHV5B.
[b] The values in parentheses indicate that incomplete ORFs were used.
[c] The divergence between EEHV5A and EEHV5B (5A-5B), EEHV5A and EEHV2 (5A-2), EEHV5B and EEHV2 (5B-2), and EEHV5A and EEHV1 (5A-1) is shown.
[d] These values show divergence between EEHV5B and EEHV1A.

11% for the outside flanking regions within the next adjacent proteins ORF-K and ORF-L (Table 4). Therefore, for both CD-II and CD-III, the measured overall DNA level difference values are much less for EEHV1A-EEHV6 than for EEHV1A-1B.

Divergence of EEHV5 proteins compared to other EEHVs. The same type of analysis for EEHV5 revealed 29 EEHV5 ORFs that all have orthologues found in the same genomic position relative to their neighbors in EEHV1 and EEHV2, including related TK, RRB, OBP, gO, vGPCR1, ORF-J, ORF-M, ORF-N(vCXCL1), ORF-O, ORF-P, ORF-K, and ORF-L(IE-like) proteins (10). Although there is no direct evidence in this case, the EEHV5 genome is assumed to have the same characteristic *Proboscivirus*-specific organization with a 40-kb inversion of core gene blocks III-II-I that created the PPF–ORF-J–gN junction in EEHV1, EEHV2, and EEHV6. Table 5 provides a summary of the DNA and protein level identity relationships (expressed as percentage differences) among all ORF segments in common between EEHV5 and EEHV2. Data for the largest 25 ORFs evaluated are listed in order across from left to right matching the gene organization determined from the overlapping phage insert map for the central 85-kb conserved segment of EEHV2 (10) as well as for EEHV1A and EEHV1B (13, 14) with equivalent genomic positional coordinates being based on homologous regions within EEHV1A(Kimba) (GenBank accession no. KC618527) (13).

The DNA level differences between EEHV5A and EEHV2 ORFs range from 11% for U49 to 24 to 28% for U33(CysR), U71(MyrTeg), U80.5(ORF-N, vCXCL1), and U85.5(ORF-K). Eight of the most conserved proteins of EEHV5A differ from the EEHV2 versions by half or less than the DNA divergence values, which range from 4.1% to 9.2%. However, for the majority of the ORFs, the protein and DNA level divergences are closely matched with 19% for U48.5(TK) being typical, but with just a few having higher protein level differences, e.g., MyrTeg, vCXCL1, and ORF-Kex3 at 27 to 34%. Even within the highly diverged U71-gM locus, EEHV5 also proved to be significantly more closely related to EEHV2 than to EEHV1. The nearest other herpesvirus matches to the EEHV5 major capsid protein (MCP), OBP, and HEL proteins are from murine CMV (MCMV), HHV7, and HHV6 with 56, 56, and 49% divergence, respectively, and for the TK protein, the best matches are to HVsaimiri and EBV at 68 and 73% divergence over just a 200-aa block.

Chimeric nature of the EEHV5A and EEHV5B genomes. Atkins et al. (17) reported that the two strains of EEHV5 found in the Texas herd showed very few differences from the prototype EEHV5(NAP28) strain within the POL codehops region. However, whereas the U71-gM locus for NAP28 and NAP58 was nearly the same, that for NAP50 differed by 4.3% (28/653 bp). Expansion of all three EEHV5 sequence data sets as reported here (Tables 1

and 2; see Table S1 in the supplemental material) revealed a mixed chimeric pattern where some loci show very little variability, but others have large differences. Overall, the first two EEHV5 strains analyzed (NAP28 and NAP50) proved to be very similar, displaying just 0.45% total nucleotide variations, but the third strain examined EEHV5(NAP58) showed major chimeric effects within just a subset of the gene loci evaluated. Therefore, following the precedent for the partially chimeric EEHV1A versus EEHV1B genomes (10), these two subgroups will be referred to as EEHV5A and EEHV5B.

The most extensive studies were carried out with EEHV5A (NAP50) and EEHV5B(NAP58) and comparative data for all ORFs analyzed in common between them are also listed in Table 5, with the nine individual hypervariable ORFs that show greater than 4.5% DNA level divergence between EEHV5A and EEHV5B being indicated in bold text. In summary, five out of the total of 12 PCR gene loci examined in common among these two EEHV5 strains showed hypervariable DNA sequence patterns. Although just a single example of an EEHV5B strain has been analyzed so far, the situation is reminiscent of that found previously between the EEHV1A and EEHV1B groups, where five characterized independent strains of EEHV1B maintain linked CD-I, CD-II, and CD-III chimeric domains compared to over 30 strains studied of EEHV1A (10). The divergence patterns observed are summarized below, and a pictorial illustration of five of the six transitional boundaries between the EEHV5A and EEHV5B chimeric domains are presented as SimPlot DNA comparison diagrams in Fig. 5. Only the right-hand chimeric boundary of CD-III is not shown, because we do not yet have data for this region in EEHV5A. However, highly diverged versions of both the Ser/Thr-rich glycoproteins ORF-P and ORF-O are present in EEHV5B here, but ORF-Q is missing or deleted, similar to the situation in EEHV2 (Table 5).

(i) EEHV5A-EHV5B (5A-5B) chimeric domain I (CD-I). The first divergent region found lies within a contiguous 4.8-kb locus encompassing both U39(gB) and U38(POL). Here, an internal 2.4-kb segment differs by 270 nucleotide polymorphisms (11%), including all except an N-terminal 500-bp segment of the glycoprotein B gene plus 330 bp from the N terminus of the adjacent DNA polymerase gene. The CD-I region also displays 5.1% differences at the protein identity level. In comparison, EEHV2 differs from both the EEHV5A and EEHV5B strains in the matching part of U39(gB) by 15% and 7% at the DNA and protein levels, respectively. The entire diverged CD-I block occupies positions 430 to 2,850 (indicated by an arrow in Fig. 5a, representing Kimba coordinates 74,428 to 76,894), in which the first 900 bp displays 20% DNA divergence, whereas the other 1,500 bp slopes from 10% down to just 5% divergence as gB crosses into POL at position 2,530.

(ii) 5A-5B chimeric domain II (CD-II). For CD-II, both the left and right transitional boundaries (indicated by an arrow at positions 900 and 2,000) are displayed across a 2.8-kb locus in Fig. 5b. Within this short 1.1-kb interval (Kimba coordinates 106,256 to 107,335), the DNA divergence between EEHV5A and EEHV5B reaches 16.5% (188 bp), almost the same as the 17.4% level between EEHV5A and EEHV2 here. The variable region encompasses contiguous N-terminal segments of both TK and U49, but outside this region neither U48(gH) nor U50 show any significant differences. Note that the EEHV5A-5B CD-II block is displaced by about 4 kb to the right compared to the location of the CD-II block in EEHV1A-1B.

(iii) 5A-5B chimeric domain III (CD-III). For CD-III, the SimPlot diagram in Fig. 5c compares just the left boundary across a 4.2-kb segment, covering six genes from U77(HEL) to ORF-O. Unlike EEHV1B-1A, there is no significant divergence between EEHV5B-5A in ORF-M until a dramatic chimeric effect begins at the sharp transition point indicated by an arrow at position 1,600, and continues for at least 2.7 kb encompassing just 300 bp from the C terminus of ORF-M, as well as all of the adjacent U80.5(ORF-N), U81(UDG), and U82(gL) genes and at least part of U82.5(ORF-O ex-3). Overall, this segment of the CD-III region of EEHV5B (Kimba coordinates 145,170 to 147,958) differs from EEHV5A by 27% (739 bp), and by 28, 27, 27, and 23% at the amino acid level for these latter four proteins. These values are all larger than for EEHV5A from EEHV2 at 23, 23, 17, and 21% (Table 5). Finally, after a gap with missing data for EEHV5A across the expected remainder of U82.5(ORF-O) plus U83.5(ORF-P) further to the right, the entire next adjacent 3.9-kb PCR locus encompassing most of U85.5(ORF-K) and the C terminus of the IE-like ORF-L gene (plus a large 500-bp 3′-intergenic domain between them) displayed just 1.1% DNA variability. Therefore, although the hypervariable EEHV5B-5A CD-III domain is at least 2.7 kb in size, it could extend further to the right, but it is unlikely to be more than 5 kb in total length, assuming that the EEHV5A ORF-P gene is present and the ORF-Q gene is absent, as in both EEHV2 and EEHV5B.

(iv) Central intervening constant regions. Both PCR loci examined within the 29 kb between CD-I to CD-II are very highly conserved, as are five of the seven loci within the 33 kb between CD-II and CD-III and all three loci outside and to the right of CD-III. However, the 825-bp U51(vGPCR) locus of EEHV5A and EEHV5B provided an exception by differing uniformly throughout its length by 4.6% at the DNA level and by 2.7% at the protein level. The only other significant variability found occurs within the U71-gM PCR locus, which displays 8.7% (24/680) nucleotide and 14% amino acid differences between the EEHV5A and EEHV5B versions, with most of the latter variability falling just within the U71(MyrTeg) gene.

In contrast to the five variable segments described above, all seven other EEHV5 gene loci evaluated in common for both EEHV5A and EEHV5B (Table 2; see Table S1 in the supplemental material) show nucleotide differences ranging from just 0.2 to 1.3% with an average of only about 0.5% between the two subtypes. This level is equivalent to that expected for strains of the same species and is also about 6-fold lower than the 3.5% value for a large segment of the nonhypervariable core genomic background between EEHV1A and EEHV1B (10). Finally, only the originally recognized U71-gM variable locus displayed a further discordant chimeric feature, with the EEHV5(NAP28) strain matching EEHV5B(NAP58) instead of having the EEHV5A (NAP50) subtype pattern. Two other examples of EEHV5 are now known; one from a lethal case in the United Kingdom designated EEHV5(EP24) (8) and EEHV5(NAP59) from a positive trunk wash fluid sample at a different U.S. facility. Both proved to have very close matches to the typical EEHV5A features at all four loci tested (data not shown).

Overall, the nucleotide differences measured between common segments of EEHV5A and EEHV5B total 1,313 out of 22,550 bp (5.8%). However, these are concentrated primarily within the five variable domains above, with just 54 nucleotide differences occurring outside these domains. In fact, the three largest variable segments, CD-I, CD-II, and CD-III, together contain 1,197 of these nucleotide differences but occupy just 6,266 bp, represent-

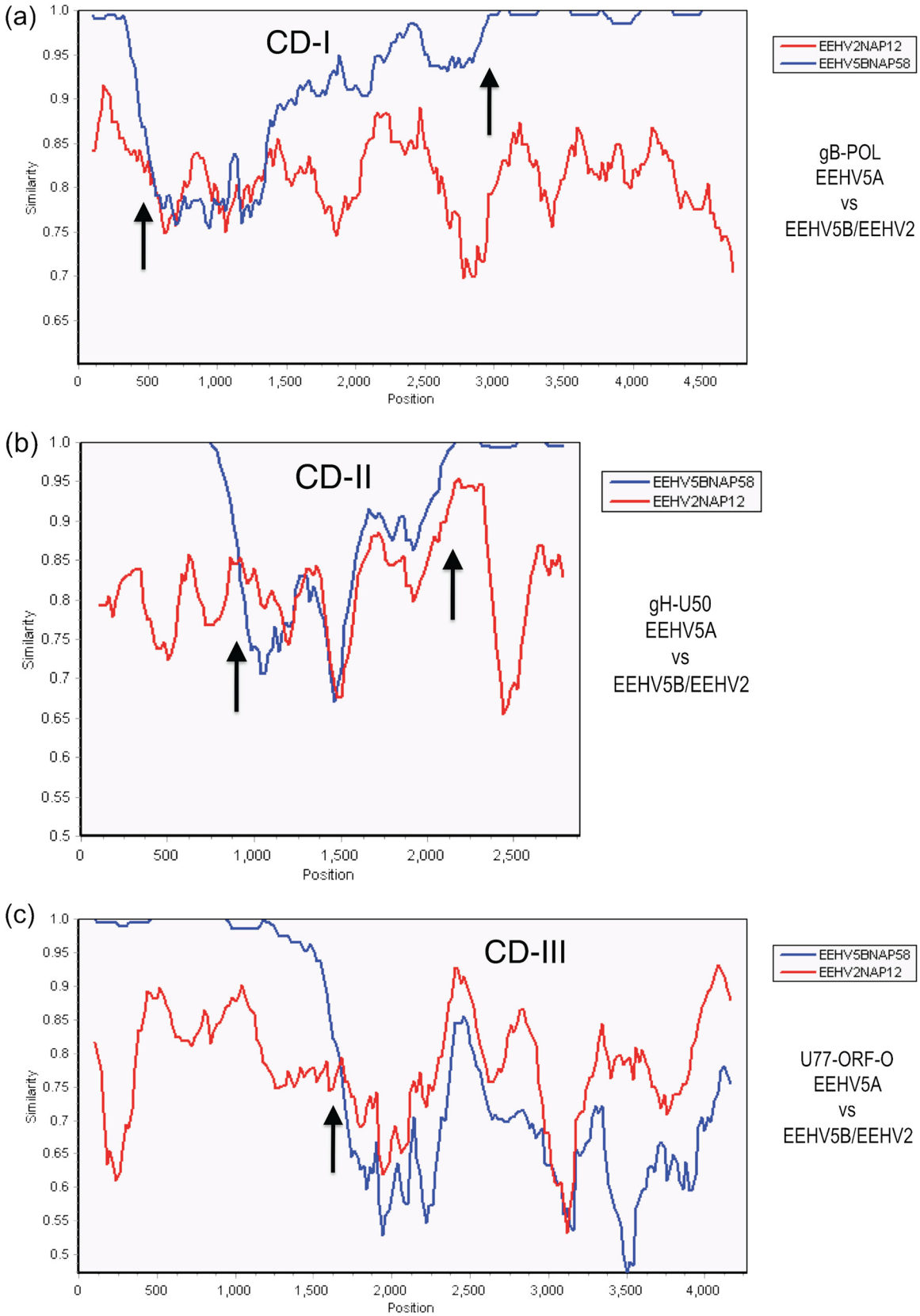


FIG 5 Evaluation of the EEHV5A-5B chimeric domain patterns and boundaries relative to EEHV2. The diagrams show SimPlot comparisons between EEHV5A, EEHV5B, and EEHV2 across all three EEHV5B chimeric domain regions, including all of CD-I (2.4 kb) and CD-II (1.1 kb) and the left-hand side boundary of CD-III (>2.7 kb). (a) CD-I. The 4,800-bp U39(gB)-U38(POL) segment from across map coordinates 74,002 to 78,861 for EEHV5A(NAP50) compared to EEHV5B(NAP58) (blue) and to EEHV2(NAP12) (red) is shown. (b) CD-II. The 2.8-kb U48(gH)–U48.5(TK)–U49–U50 segment from across map coordinates 105,396 to 108,178 for EEHV5A(NAP50) compared to EEHV5B(NAP58) (blue) and to EEHV2(NAP12) (red) is shown. (c) CD-III. The left-hand side only of the 4,300-bp segment across map coordinates 143,644 to 147,958 for EEHV5A(NAP50) encompassing U77(HEL C terminus), U77.5(ORF-M), U80.5(ORF-N), U81(UDG), U82(gL), and U82.5(ORF-Oex3) compared to EEHV5B(NAP58) (blue) and EEHV2(NAP12) (red) is shown.

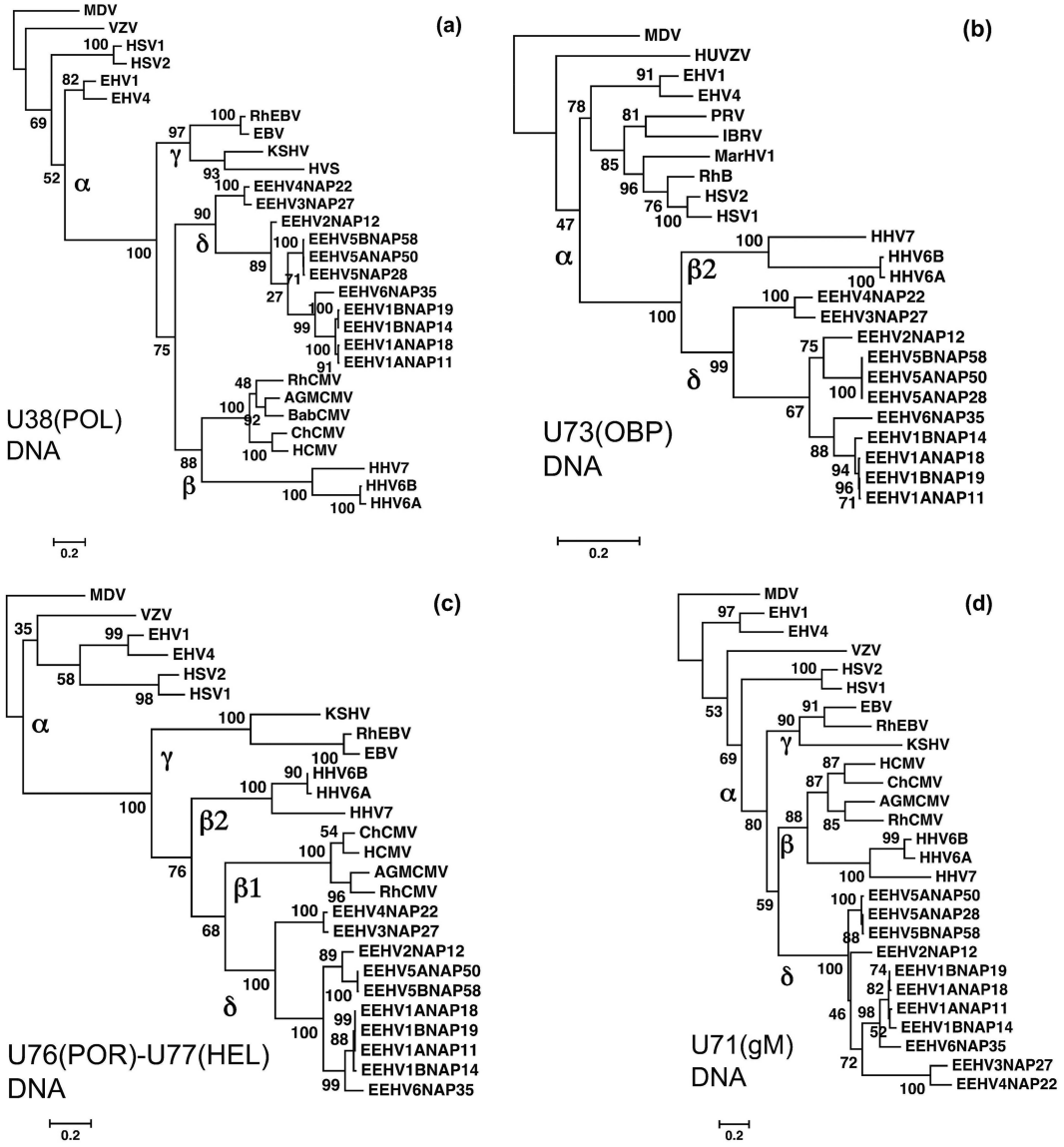


FIG 6 Linear DNA level phylogenetic trees comparing four gene loci from EEHV3, EEHV4, EEHV5A, EEHV5B, and EEHV6 with their orthologues in EEHV1, EEHV2, and other key herpesviruses. The diagrams present linear Bayesian maximum likelihood phylogenetic dendrograms at the nucleotide level for a set of representative EEHV gene loci from the five EEHV3, EEHV4, EEHV5A, EEHV5B, and EEHV6 (proposed *Deltaherpesvirinae* [δ]) genomes determined here, together with orthologues from EEHV1A, EEHV1B, and EEHV2 (10). These are compared with matching segments of selected orthologous gene loci among representative herpesviruses from all three other mammalian subfamilies, *Alphaherpesvirinae* (α), *Gammaherpesvirinae* (γ), and *Betaherpesvirinae* (β). Alternative virus names and GenBank accession numbers for both the non-EEHV and EEHV genome file sources used are listed in supplemental material or in Table S1 in the supplemental material. Bootstrap values are shown as percentages at the nodes. The final DNA segment sizes were as follows for panels a to d: (a) U38(POL) locus, DNA polymerase, 927 bp; (b) U73(OBP) locus, origin binding protein, 576 bp; (c) U76(POR)-U77(HEL) locus, portal protein plus helicase subunit; (d) U71-U72(gM) locus, myristylated tegument protein plus glycoprotein M, 353 bp. All four panels use Marek's disease alphaherpesvirus (MDV) as an outgroup.

ing 19.2% divergence. Furthermore, the DNA level differences between EEHV5A versus EEHV5B in these three blocks all reach a very substantial fraction of (and in places greater than) the genetic divergence level between the EEHV5A and EEHV2 versions. Just as for EEHV1B, the most parsimonious interpretation would seem to be simple ancient chimeric exchange events, rather than highly selective genetic drift after the initial branching event from the last common ancestor of both EEHV2 and EEHV5.

Phylogenetic tree comparisons for EEHV5A, EEHV5B, and EEHV6 relative to other probosciviruses and mammalian herpesvirus subfamilies. Linear distance-based phylogenetic tree dendrograms comparing selected EEHV5 and EEHV6 gene loci or protein segments in comparison to their orthologues in other EEHV species and within the broader context of other subfamilies and genera of the *Herpesviridae* are presented in Fig. 6 and 7. The four DNA level trees in Fig. 6a to d demonstrate the relative differences between the EEHV5 and EEHV6 versions from each other and from their closest orthologues in EEHV2 or EEHV1A and EEHV1B. In all cases, the EEHV6 plus EEHV1A and EEHV1B versions cluster together, as do the EEHV2 plus EEHV5A and EEHV5B versions. For the U38(POL) DNA locus, the EEHV1-6 and EEHV2-5 differences are closely equivalent to those measured

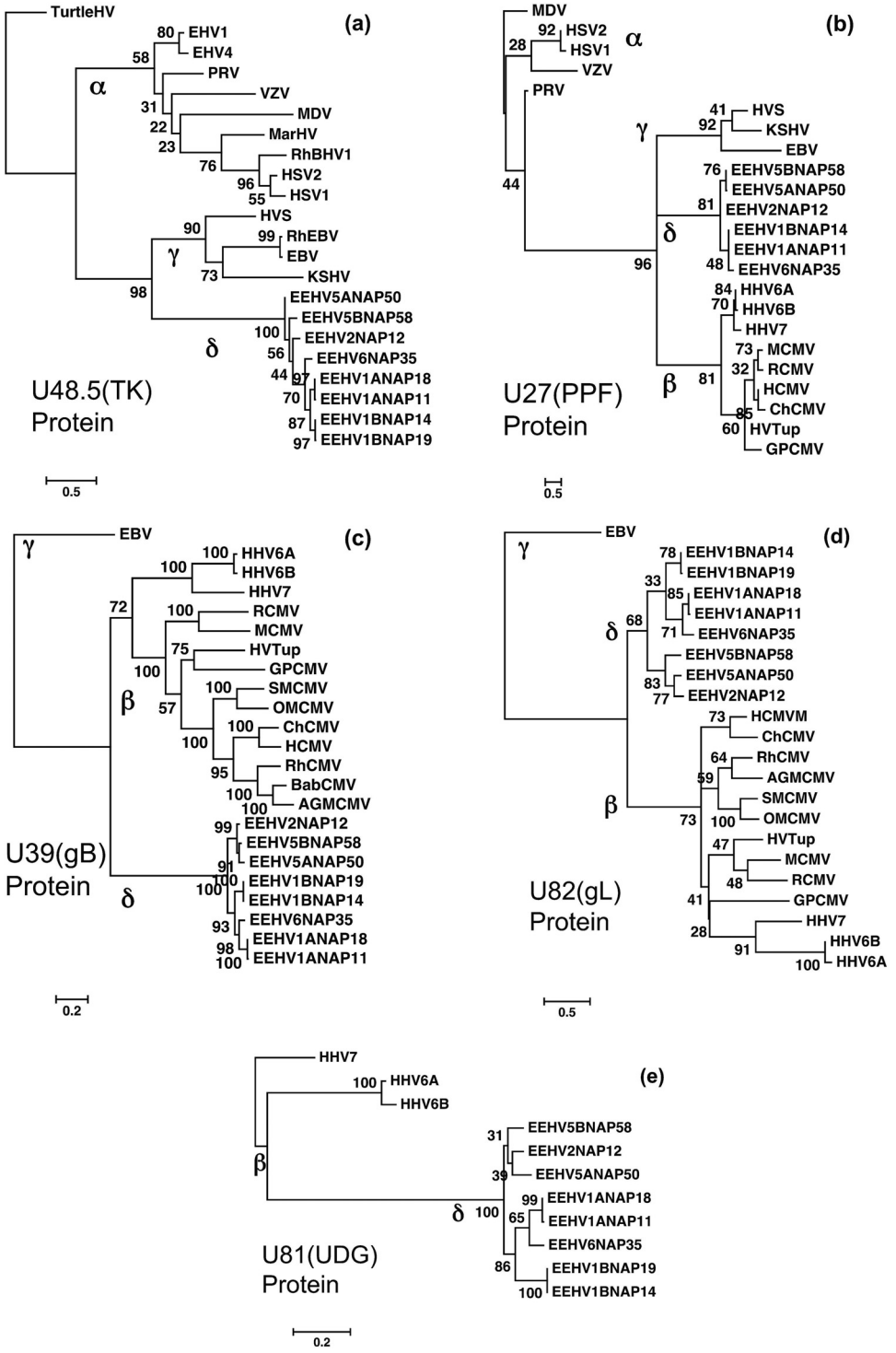


FIG 7 Linear protein level phylogenetic trees comparing five protein segments from EEHV5 and EEHV6 with their orthologues in EEHV1, EEHV2, and other key herpesviruses. The diagrams present linear Bayesian maximum likelihood phylogenetic dendrograms at the amino acid level for a set of representative proteins from the prototype EEHV5A, EEHV5B, and EEHV6 genomes, as well as up to five other orthologues from EEHV1A, EEHV1B, and EEHV2 (proposed *Deltaherpesvirinae* [δ]). These are compared to matching segments of orthologous gene loci from selected herpesviruses representative of the *Alphaherpesvirinae* (α), *Gammaherpesvirinae* (γ), or *Betaherpesvirinae* (β) subfamilies. Bootstrap values are shown as percentages. The final protein segment sizes used were as follows for panels a to e: (a) U48.5(TK), thymidine kinase, 244 aa, with turtle herpesvirus (TurtleHV) used as the outgroup; (b) U27(PPF), polymerase processivity factor, 110 aa, with MDV used as the outgroup; (c) U39(gB), glycoprotein B, 727 aa, including only the *Betaherpesvirinae* as references with EBV used as the outgroup; (d) U82(gL), glycoprotein L, 85 aa, with only *Betaherpesvirinae* included as references and EBV used as the outgroup; (e) U81(UDG), uracil DNA glycosylase, 252 aa, with only *Roseolovirus* species used for reference and HHV7 used as the outgroup.

between all five Old World primate and great ape *Cytomegalovirus* genus versions shown (Fig. 6a), and for U76(POR)-U77(HEL), the EEHV1-6 to EEHV2-5 differences are closely equivalent to those between the four CMV versions shown (Fig. 6c). However, for both, they are significantly greater than for the rhesus and human EBV, herpes simplex virus 1 (HSV1)-HSV2, and EHV1-EHV4 pairs (Fig. 6a and c). Similarly, for the U73(OBP) locus (Fig. 6b), the EEHV1-6 and EEHV2-5 differences are as large as those between the three rhesus and human *Simplexvirus* genus species shown rhesus B virus (RhBV), HSV1, and HSV2. Again, for U71-gM (Fig. 6d), the EEHV1-6 and EEHV2-5 divergence distances most closely match those between HSV1 and HSV2 and the EEHV1-5 to EEHV2-5 distances are nearly equivalent to those of the four Old World and great ape CMVs and the human and rhesus EBVs. Most importantly, in all three relevant POL, POR-HEL, and U71-gM DNA trees, the *Roseolovirus* versions cosegregate with or closer to the primate *Cytomegalovirus* branch than they do to the *Proboscivirus* clade (Fig. 6a, c, and d).

The five protein level phylogenetic tree dendrograms presented in Fig. 7 were also chosen to illustrate the patterns and levels of divergence between EEHV5A and EEHV5B compared to those for EEHV1A and EEHV1B where possible in selected variable gene loci. The U39(gB), U48.5(TK), and U27.5(PPF) protein trees (Fig. 7a, b, and c) include comparative examples from all three mammalian subfamilies, whereas to reveal more detail within the AT-rich branch of the *Proboscivirus* genus, the U82(gL) tree (Fig. 7d) includes only the betaherpesvirus versions (with EBV as the outgroup) and the U81(UDG) tree (Fig. 7e) includes only the three human examples of the *Roseolovirus* genus. The gB, gL, and UDG dendrograms (Fig. 7c, d, and e) illustrate well the much closer relatedness of the EEHV6 versions to EEHV1A than to EEHV1B for these examples from within the CD-1 and CD-III loci, whereas the gL and UDG trees show the same effect for EEHV2 and EEHV5A in relation to EEHV5B. Both the PPF and gL trees also show the divergence between EEHV1-6 and EEHV2-5 to be equivalent to that between the Old World and great ape CMVs, although the EEHVs seem to have been much more constrained within the gB protein. Again, in the three relevant gB, gL, and PPF protein trees, the *Roseolovirus* versions cosegregate with all other members of the betaherpesvirus subfamily (including tree shrew herpesvirus [HVTup], guinea pig CMV [GPCMV], MCMV, rat CMV [RCMV], and both New World and Old World primate CMVs) rather than with the *Proboscivirus* clade (Fig. 7b, c, and d).

Phylogenetic tree comparisons for the two GC-rich branch EEHVs. The DNA phylogenetic tree dendrograms in Fig. 6a to d also show genetic divergence distances across four of the PCR loci that were analyzed for EEHV3 and EEHV4. These are compared to their orthologues in the *Alphaherpesvirinae*, *Betaherpesvirinae*, and *Gammaherpesvirinae* mammalian subfamilies, including key representative examples from the *Cytomegalovirus*, *Muromegalovirus*, *Roseolovirus*, *Simplexvirus*, *Varicellovirus*, *Mardivirus*, *Lymphocryptovirus*, and *Rhadinovirus* genera. The relative branching positions of the U38(POL), U73(OBP), U76(POR)-U77(HEL), and U71-U72(gM) loci from EEHV3 and EEHV4 compared to other herpesviruses in the phylogenetic trees are highly similar to one another. In particular, they are most closely related to the other *Proboscivirus* versions, but they show significantly greater (up to twice as much) divergence from both EEHV1-6 and EEHV2-5 than EEHV1-6 and EEHV2-5 show from each other. Overall, the DNA level genetic distances observed between core genes of the GC-rich branch *Proboscivirus* genomes from those of the AT-rich branch are at least as great as those between the HSV, rhesus HSV (RhHSV), and PRV plus EHV1 alphaherpesviruses (Fig. 6a, b, and c), which are classified as different genera, as well as nearly equivalent to those between all Old World and New World primate gammaherpesviruses, including lymphocryptoviruses (EBVs) and rhadinoviruses (Kaposi's sarcoma-associated herpesvirus [KSHV] and HVS) (Fig. 6a, c, and d). They are also much greater than across the whole spectrum of Old World monkey and great ape CMVs (Fig. 6a, c, and d). Furthermore, the two major AT-rich and GC-rich EEHV branches are also at least as far diverged as the *Roseolovirus* and primate *Cytomegalovirus* versions in two of these DNA trees (Fig. 6a and d), although not so for POR-HEL (Fig. 6c).

Estimates of divergence ages by comparison with New World, Old World, and great ape versions within the alpha-, beta-, and gammaherpesvirus subfamilies. For the primary purpose of estimating the divergence ages among the different EEHV branches, Fig. 8 shows a Bayesian nearest-neighbor phylogenetic tree for the same 1,077-bp segments of the U38(POL) DNA sequence that are common among all eight EEHV types that were used in Fig. 2. However, here the protein versions are specifically compared with the matching POL protein regions from three selected sets of primate herpesviruses: those for the *Simplexvirus*, *Lymphocryptovirus* (LCV) and *Cytomegalovirus* (CMV) genera. In each case, the virus species used include representatives from human and several other great ape hosts, as well as from both Old World monkeys (baboon, rhesus, and African green) and New World monkeys (squirrel, owl, marmoset, and cebus), and in the case of the CMVs, the more primitive tree shrew herpesvirus (HVTup) as well. The primate herpesviruses were chosen here because of the relatively well established divergence dates of their mammalian host species (21), and because of previously published evaluations concluding that (with just a few specific exceptions), these viruses seem to have largely coevolved together with their natural hosts throughout primate evolution.

The generally accepted divergence dates at branch point nodes for New World monkeys from Old World monkeys of 43 million years ago (Mya), of great apes from Old World monkeys of 30 Mya, of gorillas from modern humans (9 Mya), and of chimpanzees from modern humans of 6 Mya are derived from Steiper and Young (21). Within the Old World monkey lineage, values of between 9 and 15 Mya are also usually employed for the branching of African green, baboon, and rhesus monkey hosts. Furthermore, all of these relative branching patterns match quite well for the intact DNA POL proteins among a set of eight primate CMVs that were evaluated previously by Alcendor et al. (22). However, there are two known major exceptions to linear evolutionary patterns among primate herpesviruses that are relevant to the data included here in Fig. 8. The first exception is for the human LCV EBV, which together with some chimpanzee and gorilla versions (*Pan paniscus* LCV1 [PpanLCV] and gorilla LCV1 [GorLCV1]), but unlike the orangutan and another gorilla version (*Pongo pygmaeus* LCV [PpymLCV] and GorLCV2) was evidently anciently derived from an Old World monkey LCV source (23, 24). Second, both the particular lineages of gorilla GorCMV and chimpanzee CMV (ChCMV) shown in this tree are thought to have swapped hosts at some point (25, 26), and therefore should be considered to have inverted divergence ages of 6 and 9 Mya from human CMV (HCMV).

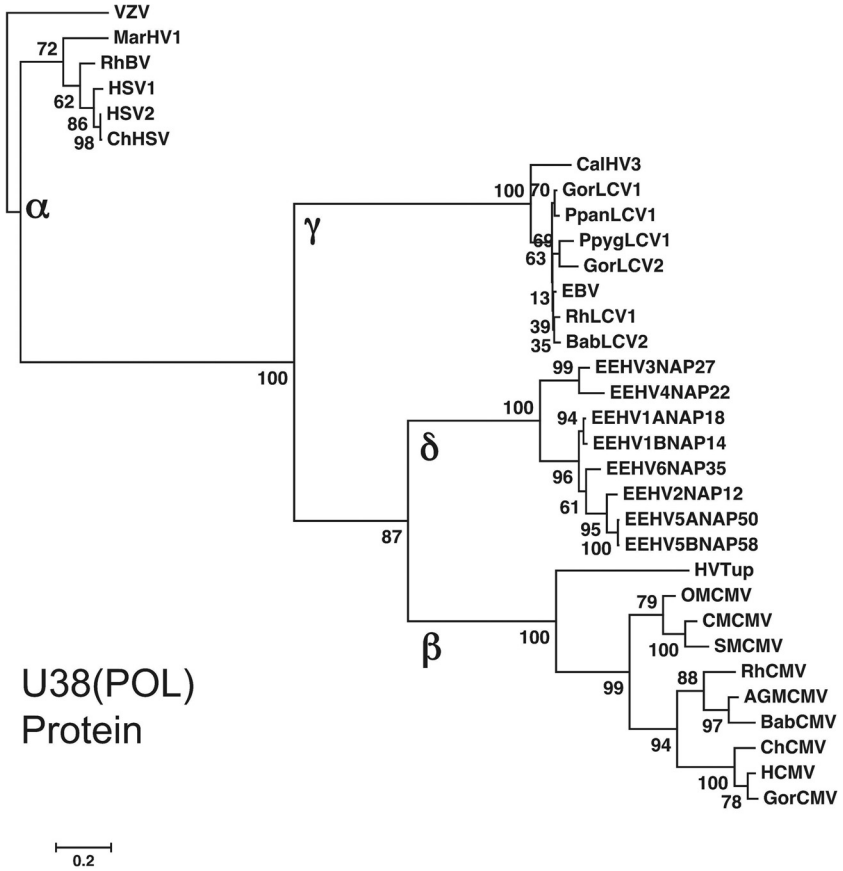
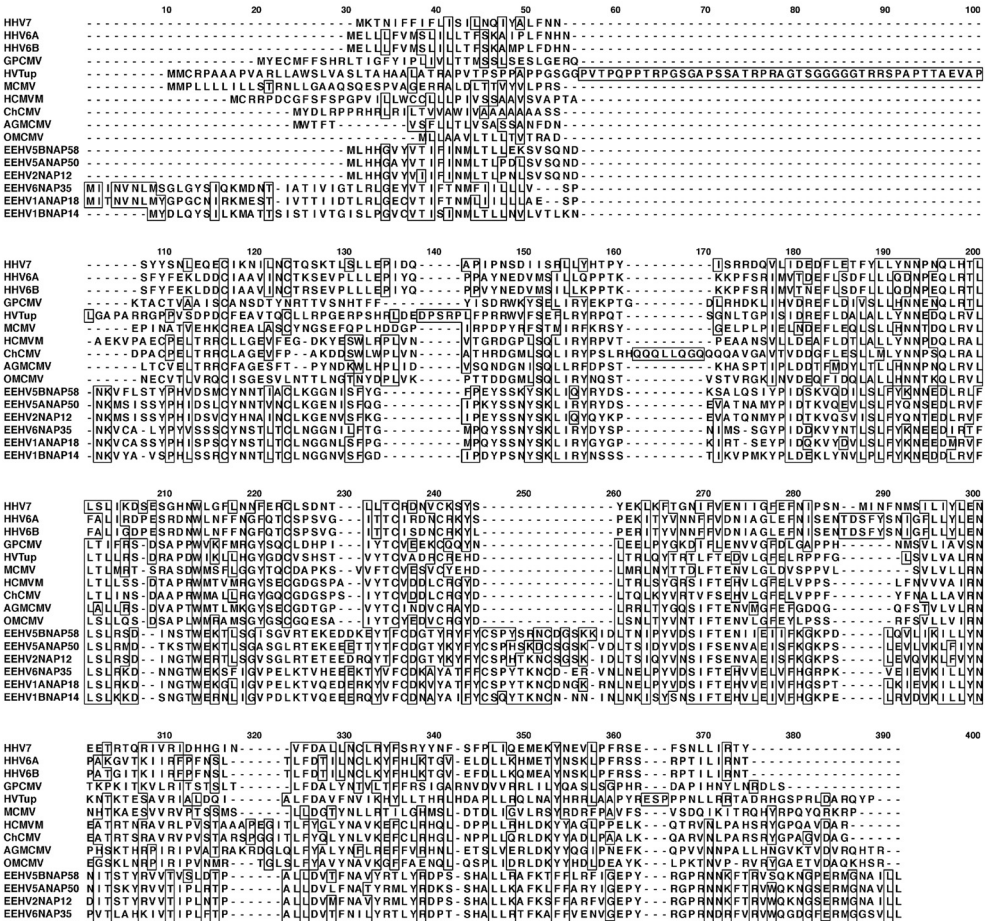


FIG 8 Comparison of EEHV POL protein divergence levels with great ape plus New World and Old World primate versions within the alpha-, beta-, and gammaherpesvirus subfamilies. The diagram shows a linear distance-based Bayesian phylogenetic tree dendrogram at the amino acid level for each of the eight prototype EEHV genomes (proposed *Deltaherpesvirinae* [δ]) across the largest segment of U38(POL) in common (1,080 bp, Kimba equivalent coordinates 77,783 to 78,863). To allow direct comparisons of estimated ages of divergence (see the text), these are compared with matching segments of orthologous gene loci from selected primate herpesviruses representative of just the *Simplexvirus* (*Alphaherpesvirinae* [α]), *Lymphocryptovirus* genus (*Gammaherpesvirinae* [γ]), and *Cytomegalovirus* (*Betaherpesvirinae* [β]) genera. Human VZV (*Varicellovirus* genus) was used as the outgroup. The evolutionary history was inferred by using the maximum likelihood method based on the Kimura two-parameter model with initial trees for the heuristic model obtained by applying the neighbor-joining method. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Evolutionary analyses were conducted in MEGA5. Bootstrap values are shown as percentages.

A direct comparison across the different clades within the tree indicates that the GC-rich and AT-rich branches of EEHV have diverged somewhat more than the New World and Old World monkey versions of primate simplexviruses and considerably more than those of lymphocryptoviruses but less so than the same two lineages within the *Cytomegalovirus* genus. Similarly, the four AT-rich branch EEHVs have diverged about equally to many of the lineages within the great ape, Old World, or New World primate versions of the cytomegaloviruses, but considerably further than those same branches within the *Simplexvirus* and *Lymphocryptovirus* genera. Evidently the rates of evolutionary drift here have varied by up to 2-fold across these three major subfamilies of primate herpesviruses. However, assuming that the EEHVs diverged at the average rate of these three other herpesvirus genera leads to extrapolated estimates of the major *Proboscivirus* branching point nodes to be about 35 to 40 Mya for the GC-rich branch from the AT-rich branch and close to 18 to 20 Mya for the EEHV1-EEHV6 lineage from the EEHV2-EEHV5 lineage within the AT-rich branch. Similarly, the split between EEHV1 from EEHV6, as well as that between EEHV2 from EEHV5, could reasonably be projected to have occurred about 12 to 14 Mya and that between EEHV3 from EEHV4 about 6 to 8 Mya.

Patterns of amino acid divergence across selected highly variable *Proboscivirus* proteins. Finally, five illustrated examples of amino acid Clustal alignments for highly variable proteins from within the AT-rich branch of the probosciviruses are presented in Fig. 9 and 10. In the first example, the 320-aa EEHV U82(gL) proteins are compared to those of a reference set of their betaherpesvirus orthologues representing both the *Cytomegalovirus* and *Roseolovirus* genera (Fig. 9a). While divergence between the proposed *Deltaherpesvirinae* and *Betaherpesvirinae* versions occurs relatively uniformly throughout the whole length of the U82(gL) proteins, the different EEHV lineages and species are especially variable in both the length and amino acid sequences across their N-terminal regions. In the second example, the 315-aa EEHV U82(UDG) proteins are compared with just those of the three human *Roseolovirus* species (Fig. 9b). Here the situation is dramatically different, with the C-terminal 200-aa domain being highly conserved in all (as are host vertebrate UDGs here as well) because of the functional enzymatic motifs. In contrast, the N-ter-

(a) U82 (gL)



(b) U81 (UDG)

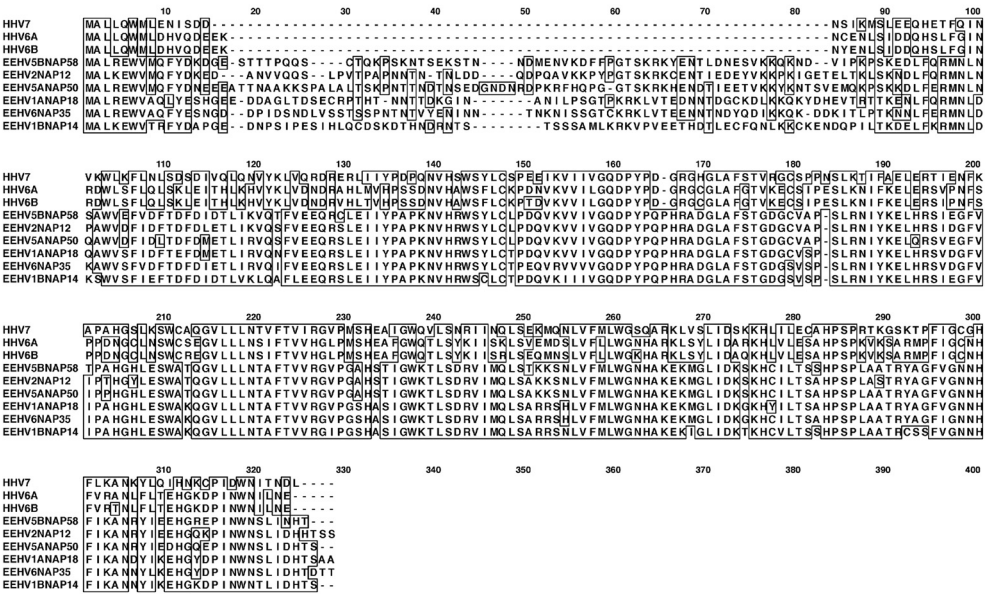


FIG 9 Clustal alignment illustrating variability patterns among the intact glycoprotein L (gL) and uracil DNA glycosylase (UDG) proteins from distinct types of EEHV genomes. Clustal-W presentations comparing the predicted primary amino acid sequence alignments of the prototypes of all six available types and subtypes of AT-rich branch *Proboscivirus* genomes compared to their closest orthologues in either all *Betaherpesvirinae* or in just the *Roseolovirus* genus. GenBank accession numbers for the EEHV proteins and betaherpesvirus reference species used are given in the supplemental material. (a) U82(gL), glycoprotein L, intact proteins; (b) U81(UDG), uracil DNA glycosylase, intact proteins. Positions with 100% similarity are boxed. Gaps introduced to maximize alignment are indicated by hyphens.

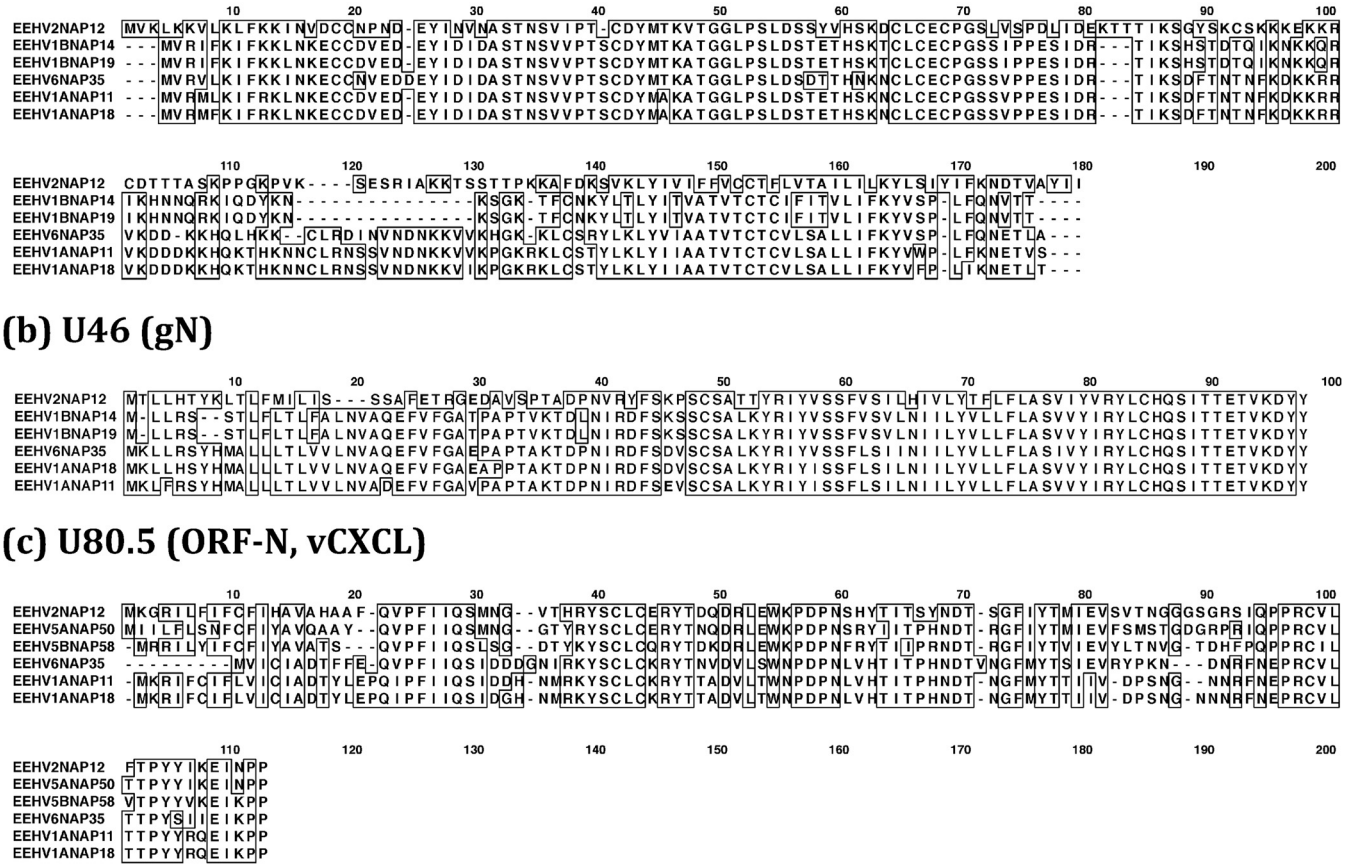


FIG 10 Clustal alignments illustrating variability patterns among the intact ORF-J, glycoprotein N (gN), and ORF-N(vCXCL1) proteins of AT-rich branch EEHV genomes. Clustal-W presentations comparing the primary amino acid sequence alignments for all available predicted *Proboscivirus* versions of three small unique or poorly conserved proteins that do not have orthologues in other herpesviruses for comparison are shown. (a) U45.7(ORF-J), intact novel proteins. (Data are not yet available for ORF-J of EEHV5A or EEHV5B.) (b) U46(gN), glycoprotein N, intact proteins. (Data are not yet available for gN of EEHV5A or EEHV5B.) (c) U80.5(vCXCL1, ORF-N), possible inhibitory alpha-chemokine ligand. [The vCXCL1(ORF-N) gene is absent from EEHV1B.] Positions with 100% similarity are boxed.

minal segments of the EEHV versions are hugely diverged both from the *Roseolovirus* versions and among themselves, including a novel inserted 65-aa block that is almost completely unrelated in each of the six distinct EEHV lineages evaluated.

The other Clustal alignments show three of the smallest and most highly diverged EEHV proteins, including ORF-J (168 aa) and the possibly inhibitory vCXCL-like protein ORF-N (96 aa) that are not known to occur in any other herpesviruses, and gN (105 aa) whose orthologues in other herpesviruses are often too diverged to be detectable in standard BLAST-P searches. For both ORF-J (Fig. 10a) and gN (Fig. 10b), only the EEHV1A, EEHV1B, EEHV6, and EEHV2 versions are available, while those for EEHV5A or EEHV5B are not yet available. Similarly, just six versions of ORF-N(vCXCL1) are included, because this gene is absent in EEHV1B (Fig. 10c). Note that both examples of ORF-J from EEHV1B display a 15-aa deletion relative to EEHV6 as well as to EEHV1A and EEHV2.

DISCUSSION

Overall classification of all EEHV types as members of the *Proboscivirus* genus and of the proposed *Deltaherpesvirinae* subfamily. The genomic sequencing data presented here provides additional evidence that Asian and African elephants harbor a large family of related EEHV species with unusual distinctive features. Combined with the results presented in the preceding accompanying paper (10), we conclude, first, that all EEHV types evaluated so far collectively form a single herpesvirus clade with genomes that seem likely to have similar and largely colinear core gene content and organizational features. Second, we conclude that even the most conserved EEHV proteins encoded in common with other herpesviruses are very highly diverged from them, with the intact versions differing by greater than 50% from their orthologues in all other herpesviruses. Third, we conclude that the EEHV1A, EEHV2, EEHV3, EEHV4, EEHV5A, and EEHV6 genomes (as well as some segments of EEHV1B and EEHV5B) are nevertheless very significantly diverged from each other in DNA sequence, although probably not much in gene content. EEHV1 itself has been classified by the International Committee on the Taxonomy of Viruses (ICTV) (27) as the first example of a new group named the *Proboscivirus* genus that was placed within the *Betaherpesvirinae* subfamily (15). Therefore, they all seem likely to be candidates for membership of the *Proboscivirus* genus.

Richman et al. (10) proposed that EEHV1A, EEHV1B, and EEHV2, the first three members of the *Proboscivirus* genus, might

be better designated as the prototypes of a newly defined *Deltaherpesvirinae* subfamily of the mammalian *Herpesviridae* family within the order *Herpesvirales*, instead of as outlier members of the *Betaherpesvirinae*. The evidence for this was based on the following: (i) the deeply branched position in both DNA and protein level phylogenetic trees of most core EEHV genes, which fall intermediate between the *Gammaherpesvirinae* and *Betaherpesvirinae* versions and do not closely align with either group; (ii) the overall gene content, including the presence of three core genes, the TK, RRB, and OBP genes, that are absent in all or most betaherpesviruses, a novel large IE-like protein (ORF-L), an alphaherpesvirus-like dyad symmetry Ori-Lyt motif, and a number of novel genes within or adjacent to the core gene block (ORF-A, -C, -J, -F, -M, -N, -O, -P, -Q, and -K); and (iii) the unique overall gene organization in which a large 40-kb segment encompassing central core gene blocks I, II, and III are inverted relative to their arrangement in all betaherpesvirus genomes. This inversion event was evidently associated with the capture or creation of several novel genes at the junctions (e.g., ORF-A, ORF-C, and ORF-J), whereas the core U45(DUT) gene, as well as the beta-specific gene block of HHV6 from U26 to U16, may all have been lost at the inversion junctions.

The same applies to all common core genes and proteins evaluated from EEHV3, EEHV4, EEHV5, and EEHV6, including the confirmed presence of the characteristic U27(PPF) to U46(gN) inversion junction in EEHV6 at least. Many EEHV5 and EEHV6 versions were included together with EEHV1A, EEHV1B, and EEHV2 within the DNA and protein level radial phylogenetic trees presented as Fig. 7, 8, and 9 in the article by Richman et al. (10). In every instance, the EEHV5 and EEHV6 versions branched together with the other EEHVs in a distinctive monophyletic cladal pattern compared to representative examples of their orthologues from all other mammalian herpesvirus subfamilies and genera. Moreover, the EEHV5A, EEHV5B, and EEHV6 genomes also all retain the two alphaherpesvirus-like core genes TK and RRB that are missing from the genomes of all betaherpesviruses, as well as the OBP gene that is also absent from both gammaherpesviruses and most betaherpesviruses except for the *Roseolovirus* genus. In addition, we identified EEHV5 and EEHV6 versions of the unique *Proboscivirus*-specific genes encoding the proteins designated ORF-M, -N, -O, -P, and -K, as well as an IE-like ORF-L protein from both, and ORF-J, gN, and gO for EEHV6. The presence of U51(vGPCR1), U73(OBP), and U71(MyrTeg) genes in EEHV5 and EEHV6 that are closely related to those in EEHV1A, EEHV1B, and EEHV2 also further strongly supports the phylogenetic evidence (Fig. 6 and 7), implying that the AT-rich branch EEHVs all fall within the provenance of the proposed *Deltaherpesvirinae* subfamily.

Differentiation of the *Proboscivirus* genus into multiple evolutionarily distinct species. While retaining many signature features found within the central core segments of EEHV1 and EEHV2, EEHV5 and EEHV6 are both relatively uniformly diverged (average 17 to 19% at the DNA level) from their closest EEHV orthologues within the AT-rich branch at all DNA loci examined. These values argue unambiguously for EEHV5 and EEHV6 having distinct species status. Interestingly, the protein level differences fall into two categories, with some having less than half the amino acid divergence values (4 to 10%) than the DNA level values, whereas most others have protein level differences (15 to 34%) that either closely match or exceed the DNA level values (Tables 4 and 5). Notably, EEHV5B (at least), similar to EEHV2, lacks the ORF-Q gene found in EEHV1A and EEHV1B. Also EEHV5A, EEHV5B, and EEHV6 all retain the vCXCL-like ORF-N gene similar to EEHV1A and EEHV2, whereas it is deleted in EEHV1B. From additional targeted PCR sequence data (not shown), EEHV2, EEHV5A, and EEHV6 also all have the highly conserved signature captured E54(vOX2-1) gene found within the novel segment of the complete EEHV1 genomes (13, 14).

One minimal yardstick that is frequently used for discriminating between genetically distinct herpesviruses is the 480-bp diagnostic DNA POL codehops region (20), although overall gene content combined with uniform differences across a number of typical well-conserved intact core proteins obviously provides superior criteria as defined in the Ninth Report of the ICTV (15, 27). In general, distinctive strains or isolates within a single species usually differ by no more than 0.2 to 1.5% within the POL codehops region, whereas the most closely related officially defined species such as HSV1 and HSV2, EHV1 and EHV4, or the human EBV plus chimpanzee and gorilla LCVs, as well as baboon, African green monkey, and rhesus LCVs, and also human and chimpanzee versions of HSV2 and HHV6 all differ there by around 3 to 6%. Even by just this simplified genetic criterion, the EEHV5 and EEHV6 POL codehops regions display 21% differences from their nearest orthologues, EEHV2 and EEHV1, respectively, suggesting that these viruses are deserving candidates for designation as individual species. However, the situation for EEHV3 and EEHV4 and with regard to the status of the EEHV1A versus EEHV1B and EEHV5A versus EEHV5B subgroups where the POL codehops region DNA differences are just 7%, 3%, and 0.5% is much more complicated.

Because of the large DNA sequence and overall G+C content divergences between the two major branches of the EEHVs, we have been able to obtain a total of only 4 to 5 kb of sequence data each spread over five noncontiguous gene loci from EEHV3 and EEHV4 directly from necropsy tissue DNA. For both genomes, all five loci proved to represent examples of a very distinctive high GC-rich branch of the *Proboscivirus* genus with between 91 to 99% G+C content in the third wobble codon position in five of the seven ORFs examined. Although hugely diverged from all four of the AT-rich branch EEHV genome types, they nevertheless retain several signature features of the gene content of the other EEHVs, especially the presence of a UL09-like U73(OBP) protein, and the same type of U71-gM locus, although there is no data available about the core domain III-II-I block inversion itself. With the most highly conserved TER protein being the only exception, the differences between the GC-rich branch and AT-rich branch orthologues for parts of the seven individual core ORFs examined here range between 30 to 50% at the DNA level and between 21 to 63% at the protein level (Table 3). Potentially, when more data becomes available, and if they should prove to contain additional subsets of unique genes, a case might also be made for the GC-rich branch to be a second genus within the *Deltaherpesvirinae* subfamily.

EEHV3 and EEHV4 themselves are considerably more closely related to each other than any other pair of EEHV types and show overall nucleotide difference levels between them of 12.5% over 3,708 bp in common. However, the POL, TER, and POR-HEL DNA loci show just 7.0, 3.8, and 3.5% nucleotide differences, respectively, whereas the less conserved U71-gM and OBP differ by

26 and 17%, respectively. Overall, we suggest that these differences, although somewhat borderline for the highly conserved TER, POR and HEL proteins, are nevertheless sufficient to justify separate species status for both EEHV3 and EEHV4, especially because of increasing evidence that the now many known additional strains of EEHV3 are endogenous to African elephants (J.-C. Zong, S.Y. Heaggans, S. Y. Long, E. M. Latimer, S. A. Nofs, M. Fouraker, V. R. Pearson, L. K. Richman, and G. S. Hayward, submitted for publication), whereas the only two known examples of EEHV4 were both found in Asian elephants (3, 7).

Apparent chimeric nature of the EEHV5B genome compared to EEHV5A. One of the three examples of EEHV5 genomes examined revealed a considerable level of localized divergence from the other two as detailed in the ORF DNA and protein level comparisons in Table 5 and in the SimPlot diagrams in Fig. 5. The total nucleotide differences in regions analyzed in common between the proposed EEHV5A and EEHV5B genome prototypes comes to 1,197 bp, but 90% of this difference is concentrated within the three discrete blocks referred to as CD-I, CD-II, and CD-III. These three loci total at least 6.2 kb in size representing more than 22% of the overall length evaluated, and they diverge by an average of 19%. Admittedly, the overall 5.6% difference observed so far between EEHV5A and EEHV5B is likely to be significantly distorted by the fact that we focused on evaluating segments equivalent to the known chimeric domains of EEHV1B. Most strikingly, the location of the mosaic genomic variability features between EEHV5B and EEHV5A rather closely resemble the patterns in at least two of the three CD regions that we described in detail previously for EEHV1B compared to EEHV1A (10). However, for EEHV1B, the overall chimeric variability is nearly twice as large as for EEHV5B, averaging 32.3% divergence distributed over three segments totaling 15 kb in size. In both cases, this unusual pattern of variability is interpreted to reflect ancient events whereby highly diverged parts of other unknown EEHV-like species have been stably recombined into the EEHV1A or EEHV5A background to create the two alternative chimeric subgroups.

Although occasional evidence for additional more modern recombinants between EEHV1A and EEHV1B has also been found, which precludes the EEHV1A and EEHV1B taxons from being given separate species status, the chimeric patterns proved to remain fully linked across all three nonadjacent CD regions in at least five independent strains of EEHV1B as well as in nearly 30 strains of EEHV1A examined so far (10). However, with only a single example of EEHV5B (and three of EEHV5A) having been identified so far, it is not yet known whether the EEHV5B chimeric segments will retain the same consistent linked pattern as found for EEHV1B. We emphasize that although the multigene chimeric domains in EEHV1A-1B and EEHV5A-5B superficially resemble the many hypervariable gene loci that are also found in extant strains of HCMV, the two situations differ significantly in that the many distinct subtype clusters of the latter are almost randomly scrambled at nonadjacent loci along different HCMV genomes, rather than displaying the highly consistent linkage patterns found for CD-I, CD-II, and CD-III in EEHV1B.

An important observation with regard to the chimeric features is that within all three major hypervariable segments, the levels of divergence between EEHV6 and EEHV1A and between EEHV5A and EEHV2 are often significantly less than the levels of divergence found between the EEHV1A and EEHV1B or the EEHV5A and EEHV5B versions. In SimPlot comparisons, the CD-I, CD-II, and CD-III regions of EEHV1B and EEHV5B both display sharp transitional boundaries indicative of chimeric crossover junctions, unlike the uniform divergence patterns found all the way across the genomes between EEHV1A-EEHV6 and EEHV5A-EEHV2 (Fig. 3, 4, and 5). Therefore, in both cases, we infer the following. (i) The chimeric segments were most likely derived simply by recombination in *trans* with one or more other novel EEHV-like genomes that had become highly diverged by standard genetic drift mechanisms, rather than having diverged at selectively high rates in *cis*. (ii) The most parsimonious scenario is that it was the EEHV1B and EEHV5B versions that acquired chimeric segments from other novel genomes, rather than EEHV1A and EEHV5A.

Estimated evolutionary divergence ages for the three sister pairs of EEHVs and the two major AT-rich and GC-rich branches of the proboscivirus. Intriguingly, at all loci examined at the DNA and protein levels, the known EEHVs group into three sister pairs with the most closely related genomes evidently being EEHV1 and EEHV6, EEHV2 and EEHV5, and EEHV3 and EEHV4 (Fig. 2, 6, 7, and 8). Importantly, the DNA divergence levels of 17% and 18% at the nucleotide level between the intact POL genes of EEHV1-6 and of EEHV2-5 are considerably higher than the 12% value for POL from human HSV1 and HSV2, which are estimated to have diverged 13 million years ago (16). They are also greater than the divergence observed in the trees between the POL genes of human EBV, and several chimpanzee, baboon, and rhesus LCVs. However, they are about equal to the DNA level differences between HCMV and ChCMV or for rhesus, baboon, and African green monkey CMVs (POL, POR-HEL, and U71-gM in Fig. 6), while being somewhat less at the protein level (PPF, gB, and gL in Fig. 7).

Chimpanzees and humans are considered to have diverged from each other about 6 million years ago (Mya), and chimpanzees and humans both diverged from gorillas 9 Mya, with the Old World primate branch that includes rhesus macaques, baboons, and African green monkeys diverging from their common ancestor with the modern great apes 30 Mya, whereas the branching between New World and Old World monkeys occurred about 43 Mya (21, 22). Curiously, for the intact POL protein, the Heberling strain of chimpanzee ChCMV (25) is noticeably much further diverged from human HCMV (16%) than are the human and chimpanzee versions of HSV2 from each other (6%), or than the human, chimpanzee, and gorilla EBVs from each other (5 to 7%). This anomaly was partially explained by the discovery of two distinct lineages each of CMV in both chimpanzee and gorilla hosts, and the recognition that the sequenced ChCMV(Heberling) strain is representative of the older gorilla-like lineage rather than the true linear chimpanzee lineage (26). However, in distance-based Bayesian phylogenetic trees, including that for primate herpesvirus POL proteins (Fig. 8), as well as those for all five other relevant EEHV genes or proteins evaluated (Fig. 6a, c, and d and Fig. 7c and d), the divergence distances of ChCMV versus HCMV proteins is similar to or even greater than those between rhesus, baboon, and green monkey SCMVs. It is also at least half or more than the distance between the great ape and Old World monkey branchpoint. Therefore, this observed relative divergence is suggestive of more like 12 to 15 Mya for the branching of ChCMV(Heberling) from HCMV.

Taking into account the anomalous status of ChCMV and assuming that the primate *Betaherpesvirinae* and *Deltaherpesvirinae*

both diverged somewhat faster in mammalian hosts than did the *Alphaherpesvirinae* and *Gammaherpesvirinae*, the POL and other phylogenetic trees (Fig. 6, 7, and 8) allowed us to make reasonable comparative estimates for the divergence dates for all of the major evolutionary branches of the *Proboscivirus* group. Thus, the projected time of divergence of 35 to 40 Mya between the GC-rich EEHV branch and the last common ancestor of all four AT-rich branch EEHVs is based on the observed nodes falling on average at an intermediate position between those for orthologous herpesviruses from Old World monkeys versus New World monkeys and between those of the great apes and Old World monkeys. Similarly, the observed branching patterns suggest that the EEHV1 plus EEHV6 branch last had a common ancestor with the EEHV2 plus EEHV5 branch close to 20 Mya, with further splits occurring more recently.

This projected age for the divergence of the major GC-rich and AT-rich branches of the *Proboscivirus* genus lies considerably beyond the 24 to 28 Mya when the mastodons diverged from the elephantids, the ancestors of all modern elephants and the woolly mammoths (28). The two major extant elephant hosts, *Loxodonta africana* and *Elephas maximas*, are themselves believed to have diverged just 7.6 Mya with *E. maximas* separating from *Mammuthus mammuthus* (the extinct wholly mammoths) about 6.7 Mya and *Loxodonta cyclotis* separating from *L. africana* about 4 Mya (29). Nevertheless, the values of 12 to 15 and 8 Mya for EEHV1-6, EEHV2-5, and EEHV3-4 branching are close enough to suggest that the existence of all three pairs of EEHV species are consistent with separate evolutionary divergence of orthologous African elephant versus Asian elephant versions within each of these branches of the *Proboscivirus* genus.

Ancient origin of elephant herpesvirus groups. Although the existence of multiple herpesviruses within the same host species may seem unusual, it is far from unprecedented, with the nine herpesviruses of humans providing an obvious example (considering HHV6A and HHV6B to be separate species). However, the human herpesviruses do include species from three different subfamilies and six different genera. In contrast, all six EEHVs (plus the EEHV1B and EEHV5B chimeric subtypes) clearly belong to just a single novel cladal group, the *Proboscivirus* genus. Based on the phylogenetic trees presented in Fig. 7, 8, and 9 of Richman et al. (10), the origin of this branch lies at a deeply diverged position about equidistant between the *Betaherpesvirinae* and *Gammaherpesvirinae* and is evidently considerably older than the split between the three currently defined genera (*Cytomegalovirus*, *Muromegalovirus*, and *Roseolovirus*) comprising the *Betaherpesvirinae*. This provides a key point supporting the concept that the EEHVs might be best designated as a novel fourth *Deltaherpesvirinae* subfamily of mammalian herpesviruses.

Collectively, the phylogenetic data also indicate that the origin of the separate herpesvirus evolutionary clade leading to the *Proboscivirus* genus was at least twice as deep as that of the Old World and New World monkey split, allowing an extrapolated estimate for this to be within the range of 80 to 100 Mya. This value compares with the suggested deepest origins of the *Herpesviridae* of 300 Mya (16) and could reasonably be judged to be close to or not long after the branching into the three currently acknowledged mammalian herpesvirus subfamilies and certainly before the branching of the *Betaherpesvirinae* and *Gammaherpesvirinae* into multiple genera. That date also correlates well with the 103 Mya (based on the short fuse mammalian molecular clock) when the *Afrotheria* (30), the ancestors of modern elephants together with their numerous extinct relatives in the mastodon and elephantid groups, apparently diverged from all other placental mammals. Therefore, it would be rather surprising to find members of the *Proboscivirus* genus infecting any other mammalian group hosts, with the plausible exception of the *Sirenia* (manatees and dugongs) and perhaps hyraxes.

Considering earlier reports describing the identification of four distinct elephant gammaherpesviruses, EGHV1, EGHV2, EGHV3, and EGHV4, in eye and genital swabs from healthy Asian and African elephants (4, 9, 23), the EEHV3 and EEHV4 genomes represent the sixth and seventh types of elephantid herpesviruses discovered. Later, a fifth novel elephant gammaherpesvirus, EGHV5, was the ninth (4), whereas EEHV5 and EEHV6 represent the overall tenth and eleventh known types of herpesviruses harbored by elephants (4). Finally, a twelfth elephant herpesvirus, EEHV7, that falls within the GC-rich branch of the *Proboscivirus* genus will be described in a subsequent paper (Zong et al., submitted). Admittedly, the fact that there are two host genera (*Loxodonta* and *Elephas*) and three extant species of modern elephants involved as potential hosts (*E. maximus*, *L. africana*, and *Loxodonta cyclotis*) is certainly a contributing factor to the unexpectedly large number of elephant herpesvirus species detected.

Natural hosts of EEHV species among wild Asian and African elephants. Acute hemorrhagic disease is an unexpected feature for herpesvirus infection of a natural endogenous host to which it should have become well adapted. Therefore, mixing of elephant host species from different continents in captive situations was originally suggested to have led to occasional cross-species primary infections to explain the disease severity (1, 31). Because EEHV1A and EEHV1B have been by far the most predominant causative agents of fatal disease in Asian elephants, the focus for this scenario has obviously fallen onto those two virus types in particular. Provocative early PCR evidence suggested that EEHV1A may be present in skin lesions in healthy wild African adult elephants (1, 32). However, we have not been able to confirm that result from further evaluation of archival African skin nodule specimens, and no other examples of EEHV1 in African elephants have been found. Furthermore, eight cases of fatal acute hemorrhagic disease in free-ranging, orphaned, and captive-reared Asian elephant calves in India have now been confirmed to have been associated with a variety of different strains of EEHV1A and a ninth case has been confirmed to have been associated with EEHV1B (6). Three more cases in range countries have also been identified as EEHV1A by DNA PCR in Cambodia or Thailand plus one as EEHV4 in Thailand (7, 33). Therefore, all three of the latter viruses, EEHV1A, EEHV1B, and EEHV4, now seem likely to be endogenous infections of Asian elephants. In addition, EEHV5 is evidently also a natural infection of Asian elephants that only rarely causes serious disease (8, 17). Questions about whether or not prior subclinical latent infections with the other subtype of EEHV1 or by EEHV5 especially might provide immunological protection against EEHV1 disease remain to be addressed in the future. In contrast, with the single exception of the Asian EEHV3 (NAP27) hemorrhagic disease case evaluated here (3), multiple examples of quiescent EEHV2, EEHV3, EEHV6, and EEHV7 have now all been found in benign lung nodules from African elephants as will be described in our subsequent paper (Zong et al., submitted).

In the wild, infection by well-adapted endogenous herpesvirus

species is usually nearly ubiquitous and occurs predominantly in an asymptomatic fashion at a very young age. In elephants, the primary infections would also be expected to be asymptomatic when transmitted by contact between young calves or from reactivated shedding in healthy adult carrier animals, especially in the presence of maternal antibodies. Delayed primary infection until after weaning in the absence of maternal antibodies, and in the absence of normal prior seroconversion to other types of these viruses, may be key factors contributing to the lethal effects of these infections seen in up to 20% of juvenile Asian elephants. The most obvious solution may be early vaccination with attenuated- or killed-virus vaccines, but this will not be feasible until the current inability to grow and propagate these viruses in cell culture conditions can be overcome. Furthermore, the many different genetically diverse forms of EEHVs now identified greatly increase the challenges involved.

ACKNOWLEDGMENTS

These studies were supported by research awards from the International Elephant Foundation and the Ringling Bros. and Barnum and Bailey Center for Elephant Conservation to L.K.R. and E.M.L. at the National Elephant Herpesvirus Laboratory at the Smithsonian's National Zoo and from National Institutes of Health research grant R01 AI24576 to G.S.H. at Johns Hopkins University. Additional research grant award funding to E.M.L., L.K.R., and G.S.H. was also provided by the Morris Animal Fund, Smithsonian Institution, and International Elephant Foundation and to G.S.H. as part of a stewardship grant from the Institute of Museum and Library Services.

We thank the many veterinary colleagues at affected zoos, circuses, and conservation programs who provided critical pathological samples for these studies, and especially Lisa Atkins, Jeff Stanton, and Paul Ling of Baylor College of Medicine for identifying and providing the high-abundance trunk wash fluid sample of EEHV5(NAP58) DNA used here. Summer student Monica Lincoln assisted in the initial PCR work identifying the second subtype of the EEHV5 TK gene.

REFERENCES

1. **Richman LK, Montali RJ, Garber RL, Kennedy MA, Lehnhardt J, Hildebrandt T, Schmitt D, Hardy D, Alcendor DJ, Hayward GS.** 1999. Novel endotheliotropic herpesviruses fatal for Asian and African elephants. Science **283:**1171–1176. http://dx.doi.org/10.1126/science.283.5405.1171.
2. **Fickel J, Richman LK, Montali R, Schaftenaar W, Goritz F, Hildebrandt TB, Pitra C.** 2001. A variant of the endotheliotropic herpesvirus in Asian elephants (Elephas maximus) in European zoos. Vet. Microbiol. **82:**103–109. http://dx.doi.org/10.1016/S0378-1135(01)00363-7.
3. **Garner MM, Helmick K, Ochsenreiter J, Richman LK, Latimer E, Wise AG, Maes RK, Kiupel M, Nordhausen RW, Zong J-C, Hayward GS.** 2009. Clinico-pathologic features of fatal disease attributed to new variants of endotheliotropic herpesviruses in two Asian elephants (Elephas maximus). Vet. Pathol. **46:**97–104. http://dx.doi.org/10.1354/vp.46-1-97.
4. **Latimer E, Zong J-C, Heaggans SY, Richman LK, Hayward GS.** 2011. Detection and evaluation of novel herpesviruses in routine and pathological samples from Asian and African elephants: identification of two new probosciviruses (EEHV5 and EEHV6) and two new gammaherpesviruses (EGHV3B and EGHV5). Vet. Microbiol. **147:**28–41. http://dx.doi.org/10.1016/j.vetmic.2010.05.042.
5. **Hayward GS.** 2012. Conservation: clarifying the risk from herpesvirus to captive Asian elephants. Vet. Rec. **170:**202–203. http://dx.doi.org/10.1136/vr.e1212.
6. **Zachariah A, Zong JC, Long SY, Latimer EM, Heaggans SY, Richman LK, Hayward GS.** 2013. Fatal herpesvirus hemorrhagic disease in wild and orphan Asian elephants in southern India. J. Wildl. Dis. **49:**381–393. http://dx.doi.org/10.7589/2012-07-193.
7. **Sripiboon S, Tankaew P, Lungka G, Thitaram C.** 2013. The occurrence of elephant endotheliotropic herpesvirus in captive Asian elephants (Elephas maximus): first case of EEHV4 in Asia. J. Zoo Wildl. Med. **44:**100–104. http://dx.doi.org/10.1638/1042-7260-44.1.100.
8. **Denk D, Stidworthy MF, Redrobe S, Latimer E, Hayward GS, Cracknell J, Claessens A, Steinbach F, McGowan S, Dastjerdi A.** 2012. Fatal elephant endotheliotropic herpesvirus type 5 infection in a captive Asian elephant. Vet. Rec. **171:**380–381. http://dx.doi.org/10.1136/vr.e6833.
9. **Wellehan JF, Johnson AJ, Childress AL, Harr KE, Isaza R.** 2008. Six novel gammaherpesviruses of Afrotheria provide insight into the early divergence of the Gammaherpesvirinae. Vet. Microbiol. **127:**249–257. http://dx.doi.org/10.1016/j.vetmic.2007.08.024.
10. **Richman LK, Zong J-C, Latimer EM, Lock J, Fleischer R, Heaggans SY, Hayward GS.** 2014. Elephant endotheliotropic herpesviruses EEHV1A, EEHV1B, and EEHV2 from cases of hemorrhagic disease are highly diverged from other mammalian herpesviruses and may form a new subfamily. J. Virol. **88:**13523–13546. http://dx.doi.org/10.1128/JVI.01673-14.
11. **Ehlers B, Dural G, Marschall M, Schregel V, Goltz M, Hentschke J.** 2006. Endotheliotropic elephant herpesvirus, the first betaherpesvirus with a thymidine kinase gene. J. Gen. Virol. **87:**2781–2789. http://dx.doi.org/10.1099/vir.0.81977-0.
12. **Richman LK.** 2003. Pathological and molecular aspects of fatal endotheliotropic herpesviruses of elephants. Ph.D. thesis. The Johns Hopkins University, Baltimore, MD.
13. **Ling PD, Reid JG, Qin X, Muzny DM, Gibbs R, Petrosino J, Peng R, Zong J-C, Heaggans SY, Hayward GS.** 2013. Complete genome sequence of elephant endotheliotropic herpesvirus 1A. Genome Announc. **1**(2): e0010613. http://dx.doi.org/10.1128/genomeA.00106-13.
14. **Wilkie GS, Davison AJ, Watson M, Kerr K, Sanderson S, Bouts T, Steinbach F, Dastjerdi A.** 2013. Complete genome sequences of elephant endotheliotropic herpesviruses 1A and 1B determined directly from fatal cases. J. Virol. **87:**6700–6712. http://dx.doi.org/10.1128/JVI.00655-13.
15. **Davison AJ, Eberle R, Ehlers B, Hayward GS, McGeoch DJ, Minson AC, Pellett PE, Roizman B, Studdert MJ, Thiry E.** 2009. The order Herpesvirales. Arch. Virol. **154:**171–177. http://dx.doi.org/10.1007/s00705-008-0278-4.
16. **McGeoch DJ, Rixon FJ, Davison AJ.** 2006. Topics in herpesvirus genomics and evolution. Virus Res. **117:**90–104. http://dx.doi.org/10.1016/j.virusres.2006.01.002.
17. **Atkins L, Zong J-C, Tan J, Mejia A, Heaggans SY, Nofs SA, Stanton JJ, Flanagan JP, Howard L, Latimer E, Stevens MR, Hoffman DS, Hayward GS, Ling PD.** 2013. EEHV-5, a newly recognized elephant herpesvirus associated with clinical and subclinical infections in captive Asian elephants (Elephas maximus). J. Zoo Wildl. Med. **44:**136–143. http://dx.doi.org/10.1638/1042-7260-44.1.136.
18. **Stanton JJ, Zong J-C, Latimer E, Tan J, Herron A, Hayward GS, Ling PD.** 2010. Detection of pathogenic elephant endotheliotropic herpesvirus in routine trunk washes from healthy adult Asian elephants (Elephas maximus) by use of a real-time quantitative polymerase chain reaction assay. Am. J. Vet. Res. **71:**925–933. http://dx.doi.org/10.2460/ajvr.71.8.925.
19. **Stanton JJ, Zong J-C, Eng C, Howard L, Flanagan J, Stevens S, Schmitt D, Wiedner E, Graham D, Junge R, Weber MA, Fischer M, Mejia A, Tan J, Latimer E, Herron A, Hayward GS, Ling PD.** 2013. Kinetics of viral loads and genotypic analysis of elephant endotheliotropic herpesvirus-1 infection in captive Asian elephants (Elephas maximus). J. Zoo Wildl. Med. **44:**42–54. http://dx.doi.org/10.1638/1042-7260-44.1.42.
20. **Rose TM.** 2005. CODEHOP-mediated PCR - a powerful technique for the identification and characterization of viral genomes. Virol. J. **2:**20. http://dx.doi.org/10.1186/1743-422X-2-20.
21. **Steiper ME, Young NM.** 2006. Primate molecular divergence dates. Mol. Phylogenet. Evol. **41:**384–394. http://dx.doi.org/10.1016/j.ympev.2006.05.021.
22. **Alcendor DJ, Zong JC, Dolan A, Gatherer D, Davison AJ, Hayward GS.** 2009. Patterns of divergence in the vCXCL and vGPCR gene clusters in primate cytomegalovirus genomes. Virology **395:**21–32. http://dx.doi.org/10.1016/j.virol.2009.09.002.
23. **Ehlers B, Dural G, Yasmum N, Lembo T, de Thoisy B, Ryser-Degiorgis MP, Ulrich RG, McGeoch DJ.** 2008. Novel mammalian herpesviruses and lineages within the Gammaherpesvirinae: cospeciation and interspecies transfer. J. Virol. **82:**3509–3516. http://dx.doi.org/10.1128/JVI.02646-07.
24. **Ehlers B, Spiess K, Leendertz F, Peeters M, Boesch C, Gatherer D, McGeoch DJ.** 2010. Lymphocryptovirus phylogeny and the origins of Epstein-Barr virus. J. Gen. Virol. **91:**630–642. http://dx.doi.org/10.1099/vir.0.017251-0.
25. **Davison AJ, Dolan A, Akter P, Addison C, Dargan DJ, Alcendor DJ,**

McGeoch DJ, Hayward GS. 2003. The human cytomegalovirus genome revisited: comparison with the chimpanzee cytomegalovirus genome. J. Gen. Virol. **84:**17–28. http://dx.doi.org/10.1099/vir.0.18606-0.

26. **Leendertz FH, Deckers M, Schempp W, Lankester F, Boesch C, Mugisha L, Dolan A, Gatherer D, McGeoch DJ, Ehlers B.** 2009. Novel cytomegaloviruses in free-ranging and captive great apes: phylogenetic evidence for bidirectional horizontal transmission. J. Gen. Virol. **90:**2386–2394. http://dx.doi.org/10.1099/vir.0.011866-0.
27. **King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (ed).** 2012. Virus taxonomy. Ninth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, CA.
28. **Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, Teeling E, Ryder OA, Stanhope MJ, de Jong WW, Springer MS.** 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science **294:**2348–2351. http://dx.doi.org/10.1126/science.1067179.
29. **Rohland N, Malaspinas AS, Pollack JL, Slatkin M, Matheus P, Hofreiter M.** 2007. Proboscidean mitogenomics: chronology and mode of elephant evolution using mastodon as outgroup. PLoS Biol. **5:**e207. http://dx.doi.org/10.1371/journal.pbio.0050207.
30. **Murphy WJ, Pringle TH, Crider TA, Springer MS, Miller W.** 2007. Using genomic data to unravel the root of the placental mammal phylogeny. Genome Res. **17:**413–421. http://dx.doi.org/10.1101/gr.5918807.
31. **Richman LK, Montali RJ, Hayward GS.** 2000. Review of a newly recognized disease of elephants caused by endotheliotropic herpesviruses. Zoo Biol. **19:**383–392. http://dx.doi.org/10.1002/1098-2361(2000)19:5<383::AID-ZOO8>3.0.CO;2-X.
32. **Richman LK, Montali RJ, Cambre RC, Schmitt D, Hardy D, Hildbrandt T, Bengis RG, Hamzeh FM, Shahkolahi A, Hayward GS.** 2000. Clinical and pathological findings of a newly recognized disease of elephants caused by endotheliotropic herpesviruses. J. Wildl. Dis. **36:**1–12. http://dx.doi.org/10.7589/0090-3558-36.1.1.
33. **Reid CE, Hildebrandt TB, Marx N, Hunt M, Thy N, Reynes JM, Schaftenaar W, Fickel J.** 2006. Endotheliotropic elephant herpes virus (EEHV) infection. The first PCR-confirmed fatal case in Asia. Vet. Q. **28:**61–64. http://dx.doi.org/10.1080/01652176.2006.9695209.