Originally published as:

**Ehlers, B., Spieß, K., Leendertz, F., Peeters, M., Boesch, C., Gatherer, D., McGeoch, D.J.
Lymphocryptovirus phylogeny and the origins of Epstein-Barr virus
(2010) Journal of General Virology, 91 (3), pp. 630-642.**

DOI: 10.1099/vir.0.017251-0

Title:

Lymphocryptovirus phylogeny and the origins of Epstein-Barr virus

Authors: Bernhard Ehlers (1*), Katja Spieß (1), Fabian Leendertz (2), Martine Peeters (3), Christophe Boesch (4), Derek Gatherer (5) and Duncan J. McGeoch (5)

(1) Research Group "Molecular Genetics and Epidemiology of Herpesviruses", Robert Koch-Institut, D-13353 Berlin, Germany
(2) Research Group "Emerging Zoonoses", Robert Koch-Institut, D-13353 Berlin, Germany
(3) UMR 145, Institut de Recherche pour le Développement, and University of Montpellier 1, Montpellier, France
(4) Max Planck Institute for Evolutionary Anthropology, Department of Primatology, Deutscher Platz 6, D-04103 Leipzig, Germany
(5) Medical Research Council Virology Unit, Institute of Virology, University of Glasgow, Church Street, Glasgow G11 5JR, United Kingdom

* Corresponding author:
Dr. B. Ehlers, Robert Koch-Institut, Nordufer 20, D-13353 Berlin
Phone: ++4918887542347; Fax: ++4918887542598; email: ehlersb@rki.de

Words: 5363 Abstract: 242

Running title: *Lymphocryptovirus* phylogeny

Abstract

Specimens from wild and captive primates were collected, and searched for new members of the genus *Lymphocryptovirus* (subfamily *Gammaherpesvirinae*) utilising PCR for the DNA polymerase gene. Twenty-one new viruses were detected. Together with previous findings, more than 50 distinct lymphocryptoviruses (LCVs) are now known, with hosts from six primate families (*Hominidae*, *Hylobatidae*, *Cercopithecidae*, *Atelidae*, *Cebidae*, *Pitheciidae*). Further work extended genomic sequences for 25 LCVs to 3.4-7.4 kbp. Phylogenetic trees were constructed, based on alignments of protein sequences inferred from the LCV genomic data. The LCVs fell into three major clades: Clade A, comprising New World viruses; Clade B, containing both Old World monkey viruses and hominoid viruses including Epstein-Barr virus (EBV); and Clade C, containing other hominoid viruses. By comparison with the primate tree, it was proposed that major elements of the LCV tree represented synchronous evolution with host lineages, with the earliest node in both trees being separation of Old and New World lines, but that some virus lineages originated by interspecies transfer. From comparisons of branch lengths, it was inferred that evolutionary substitution in Clade B has proceeded more slowly than elsewhere in the LCV tree. It was estimated that in Clade B a subclade containing EBV, a gorilla virus and two chimpanzee viruses derived from an Old World monkey LCV line approximately twelve million years ago, and another subclade containing an orang utan virus and a gibbon virus derived from a macaque LCV line approximately 1.2 million years ago.

Introduction

This paper is concerned with relationships among viruses of the *Lymphocryptovirus* genus (subfamily *Gammaherpesvirinae*, family *Herpesviridae*, order *Herpesvirales*; Davison et al., 2009). Lymphocryptoviruses (LCVs) all have primate hosts, and Epstein-Barr virus (EBV) is the single known human LCV. Historically the *Gammaherpesvirinae* subfamily was treated as consisting of the gamma-1 and gamma-2 groups. The term gamma-1 group is now superseded in formal taxonomy by *Lymphocryptovirus*. Three genera are presently assigned to the gamma-2 group (*Rhadinovirus*, *Percavirus* and *Macavirus*) and there are additional gamma-2 lineages for which the taxonomy is not yet developed (Davison et al., 2009; Ehlers et al., 2008). Phylogenetic trees for the *Herpesviridae*, based on molecular sequences, display large scale features within each of the three subfamilies that have been interpreted as showing synchronous development of major viral lineages with the lineages of the mammalian hosts, and such cospeciation has evidently been a prominent mode in the evolution of this virus family (McGeoch et al., 2006, 2008). It is considered that in the *Gammaherpesvirinae* the gamma-1 branch originated cospeciationally with the primate lineage (McGeoch et al., 2006).

EBV was the first gammaherpesvirus identified (Epstein et al., 1964), and was later classified as the type species of the genus *Lymphocryptovirus*. EBV causes infectious mononucleosis and is associated with various tumours in humans (Pagano, 1999). In Old World nonhuman primates, evidence for EBV-like LCVs was initially obtained by serological cross-reactivity to EBV, in chimpanzees (Landon et al., 1980), orangutans (Rasheed et al., 1977), gorillas (Neubauer et al., 1979), baboons (Vasiljeva et al., 1974), and diverse macaque species (Fujimoto et al., 1990; Hayashi et al., 1999; Rangan et al., 1986; Rivadeneira et al., 1999). More recently, PCR based methods have been used to detect LCVs in Old World and New World primates (Ramer et al., 2000; Cho et al., 2001; Ehlers et al., 2003; Prepens et al., 2007), and to date about 30 different LCVs are known.

In a previous study, we amplified partial sequences for DNA polymerase (DPOL; BALF5 in EBV) genes of 26 novel LCVs and analysed them phylogenetically. Three major clusters of LCVs were separated in the phylogenetic tree, one comprising LCVs of New World monkeys, the other two comprising LCVs of hominoids. One of them (Genogroup 1) was highly populated and contained beside hominoid LCVs, including EBV, also those of several Old World monkeys. The other group (Genogroup 2) was small and contained viruses of apes, but no monkey LCVs. Gorilla viruses were present in both groups. From these findings (in particular the last) we proposed that two lineages of Old World primate LCVs might exist, each with representatives in most primate species (Ehlers et al., 2003).

In the present study, we followed two experimental lines to reassess this hypothesis. First, we searched for additional LCVs of Genogroup 2, particularly in chimpanzees and Old World monkeys. Second, we set out to extend the short DPOL sequences, which were the only available

data for most of the described LCVs, by a bigenic PCR approach targeting also the glycoprotein B (gB; BALF4 in EBV) gene. For a subset of these viruses, we determined sequences that spanned up to four genes. With this much larger data set we have performed phylogenetic analyses, and have revised our interpretation of LCV evolution.

Results

Multigenic *de novo* detection of novel LCV

Over six years, blood, tissue and faeces samples (n=502) were collected from live or deceased individuals of 49 primate species (apes, Old World monkeys, New World monkeys and prosimians) as described in the *Methods* section. LCV DPOL gene sequences were successfully amplified from 50% of the samples (n=251) with pan-herpes DPOL PCR. Sequences identical to some of these were detected previously (Ehlers et al., 2003), but most came from 21 novel putative LCVs. They were listed with their names, abbreviations, hosts and GenBank accession numbers in Table 2. Combined with the 26 species detected previously (Ehlers et al., 2003), a total of 47 LCVs were discovered. The LCV-positive primate hosts originated from seven African, Asian and European countries, and were members of six different primate families, three of the *Catarrhini* (*Hominidae*, *Hylobatidae*, *Cercopithecidae*) and three of the *Platyrrhini* (*Atelidae*, *Cebidae*, *Pitheciidae*). No prosimian LCVs were detected.

For amplification of the major DNA binding protein (MDBP; BALF2 in EBV) gene and gB gene sequences, nested deg/dI primer sets (BALF2A and BALF4A, respectively) were derived from known LCV sequences (CalHV-3 BALF2 and PtroLCV-1 BALF4) and the primer binding sites placed within regions of high gammaherpesvirus conservation. With BALF2A and BALF4A sets we amplified MDBP and gB gene sequences from most of the 47 LCV species. Amplification of GgorLCV-2 and SsynLCV-1 with BALF2A and BALF4A yielded incorrect sequences or failed, since many GgorLCV-2 and all SsynLCV-1 positive samples were double-infected with GgorLCV-1 and SsynLCV-2, respectively. We therefore used PpygLCV-1 sequences – once determined - to design the alternative primer sets BALF2B and BALF4B. For 25 LCVs, the gB sequences could be connected to the corresponding DPOL sequences with Long-Distance (LD)-PCR using virus-specific gB-sense primers and DPOL-antisense primers. Contiguous sequences of 3.4 kbp spanning the 3´-part of the gB gene and the 5´-part of the DPOL gene were obtained (Table 2). For eleven out of 25 LCVs, MDBP sequences could be connected to gB with a second LD-PCR, and contiguous sequences of up to 7.5 kbp were determined, spanning two-thirds of the DPOL gene and the complete gB gene, and in some cases also the complete BALF3 gene (Figure 1).

Preliminary phylogenetic analysis showed that of the novel LCVs found in the present study, only one belonged to Genogroup 2 (HmueLCV-1) with all others in either Genogroup 1 or a clade

of New World monkey viruses. To search further for LCVs belonging to Genogroup 2, we designed a set of nested deg/dI primers based on the gB gene of GgorLCV-2 and therefore biased towards detection of Genogroup 2 LCVs (primer set BALF4C). With this set, we tested samples of chimpanzees and macaques, with gorilla samples as controls. Variants of the gB genes of GgorLCV-2 and PtroLCV-1 were detected but no chimpanzee or macaque LCV members of Genogroup 2. In addition, we designed a set of nested deg/dI primers based on the DPOL gene of GgorLCV-2 and with a 3´-base that was present in all DPOL gene sequences of Genogroup 2, but not of Genogroup 1 (primers not listed). With this set, the LCVs of Genogroup 2 were detected (in gorillas, orang utans and gibbons), but chimpanzee and macaque specimens tested negative.

Phylogenetic relationships in the genus *Lymphocryptovirus*

The 25 LCV gB to DPOL sequences determined in this study plus the homologous sequences of EBV, CalHV-3 and CeHV-15 were subjected to phylogenetic analysis. First, a phylogenetic tree based on gB sequences of 58 viruses from all three subfamilies of the *Herpesviridae* was constructed, and is shown in summary form in Fig. 2a, to provide an overview of the context of the gamma-1 lineage. This tree shows that all the LCVs fall into three major clades (designated A, B and C), and that the gamma-2 group can be utilised as an outgroup of related species to place the root of the gamma-1 tree. A tree for the BALF3 gene computed with Bayesian analysis utilising Monte Carlo Markov chains (BMCMC), and based on an alignment of 634 amino acid residues for twelve LCVs, displayed the same three clades (not shown). Fig. 2b shows a detailed gamma-1 tree, based on a 946 residue alignment of gB and DPOL amino acid sequences for 28 LCVs plus 11 gamma-2 primate viruses, and derived by BMCMC. The data for this tree represent the largest set of sequences and longest alignment available to give a robustly rooted gamma-1 phylogeny. The three major clades and their contents are labelled. Clade A contains all the LCVs of New World monkeys and no other viruses. Clade B contains all the LCVs of Old World monkeys plus some viruses of hominoids, including EBV. Clade C contains only hominoid viruses. Clades B and C correspond to Genogroups 1 and 2 respectively. On the basis of the clade contents and branching pattern we hypothesized that at the level of these major clades, the tree structure reflects synchronicity with that of the host lineages, in that divergence of New World and Old World lineages was the earliest branching event for both hosts and viruses, followed by divergence of Old World monkey and hominoid lineages. This interpretation is straightforward for Clades A and C, corresponding to New World monkey and hominoid hosts respectively, while Clade B was taken to correspond to Old World monkey hosts, leaving aside at this point the issue of hominoid viruses in Clade B.

The extent of correspondences between host and LCV lineages was then examined in detail. Fig. 3a displays a primate phylogenetic tree containing host species for all the LCVs whose sequences were included in our analyses and Fig. 3b shows a molecular clock version of the LCV

159 tree of Fig. 2b. The data for Fig. 3a were extracted from the large study of mammalian phylogeny
160 of Bininda-Emonds et al. (2007), and the figure shows the timescale derived by those authors. The
161 two trees in Fig. 3 were drawn to have the same size across the page from root to branch tips in
162 order to facilitate comparisons between them, taking divergence of Old and New World primate
163 LCV lineages as synchronous with divergence of the equivalent host lineages. In Clade C the
164 arrangement of branches for LCVs of gibbon species, orang utan (*P. pygmaeus*) and gorilla (*G.*
165 *gorilla*) matches with the host tree, as do the depths of branches except for those of the two gibbon
166 viruses. In Clade A there are four deep branches, three of which match arrangement and
167 approximate depth with the host tree while that for CalHV-3 and CpenLCV-1 is incongruent with
168 the corresponding host locus. Overall, then, Clades A and C, together with the deeper portion of
169 the LCV tree, present as mostly consistent with cospeciational development in terms of patterns
170 and proportions of branches.
171 Clade B is more complicated, in several respects: it is the most populous of the three major
172 clades; the overall depth of branch lengths is less than in Clades A and C; some loci are poorly or
173 incompletely resolved; and hominoid viruses appear at two loci among the majority Old World
174 monkey viruses. In all, Clade B presented the main challenges for interpretation of the LCV tree. A
175 tree specific to the 28 LCVs was therefore constructed, to provide the best achievable resolution of
176 Clade B with the available data. This utilised a 1042 residue alignment of gB and DPOL amino acid
177 sequences and was inferred by BMCMC. Fig. 4a shows only Clade B from this tree, with the root
178 position provided by Clades A and C, and with branch lengths drawn at a larger scale than in the
179 earlier figures. Four nodes in the tree presented in Fig. 4a have low posterior probabilities, and
180 these were reduced to multifurcations before computing a molecular clock version. Two additional
181 LCVs (CateLCV-1 and SsynLCV-2), for which less extensive sequence data were available, were
182 then interpolated into the molecular clock tree, as shown in Fig. 4b. For HmueLCV-1 (Clade C) and
183 17 Clade B viruses only short sequences of 175 bp to 430 bp were available, and these were not
184 included in the molecular clock tree. For discussion, seven clades comprising the tree were
185 designated as B1 to B7, and the multifurcated node from which five of these clades descend was
186 designated the ‘major multifurcation’ (MMF).
187 Considering only the monkey viruses in Clade B in the first instance, some aspects of the
188 branching pattern can be seen to correspond with that of the host tree, while others do not. Thus,
189 the branching relationships among SentLCV-1, EpatLCV-1, the two PhamLCVs (subclade B5) and
190 macaque LCVs in subclade B6 are congruent with the host tree. Short branch lengths and
191 branching uncertainties among the macaque LCVs limit the detail of comparisons for this grouping.
192 The loci of MsphLCV-1, CgueLCV-1, PbadLCV-1 and CnegLCV-1 do not fit into this cospeciational
193 scheme. Turning to the six hominoid viruses in Clade B, we note that these occur at two distinct
194 loci. Subclade B3 (comprising EBV, GgorLCV-1, PpanLCV-1 and PtroLCV-1) originates in the
195 midst of the monkey LCV lineages, and branching pattern within this grouping is partly but not

completely compatible with cospeciational development (in agreement with the earlier observations of Gerner et al., 2004). PpygLCV-2 and SsynLCV-2 appear together within the predominantly macaque virus B6 subclade, and are readily rationalised as late transfers from a macaque host. With respect to the possible Old World monkey cospeciational components of Clade B noted above, the branch lengths in Figs 3b and 4b are markedly shorter than would be expected from comparison with those in Fig. 3 of Clades A and C, and of the host tree, so that a cospeciational rationale for Clade B would imply that lineages in that clade have been changing more slowly than those in Clades A and C. The most straightforward way to assign a local cospeciational timeframe for Clade B is to take the node at which the SentLCV-1 lineage (B1 subclade) diverges from the other subclades as corresponding to that in the host tree (Fig. 3a) at which the lineage of the host species for SentLCV-1, *S. entellus*, diverges from lineages leading to *Macaca*, *Papio*, etc. These nodes are the earliest in Clade B and in the Old World monkey clade, respectively. The timescale shown in Fig. 4b is based on this assignment, and represents a relative substitution rate of 0.6 times that proposed for the whole gamma-1 tree in Fig. 3b.

The MMF node is a central feature of the tree in Fig. 4b, and was also observed in trees based on subsets of the available amino acid sequences and in trees for Old World primate LCVs based on DNA sequences (not shown). We take the MMF feature to be a result of the available sequence data being inadequate to resolve several closely spaced nodes. We note that the posterior probability support for subclade B2 branching from an earlier node than the MMF, as presented in Fig. 4b, is marginal. If we consider just the monkey virus lineages descending from the MMF, evidently both the *Papio* LCV line (B5) and the *Macaca* LCV line (in B6) are in cospeciation compatible loci, and the locus of the CateLCV-1 line (B7) could result from lack of resolution for placing that virus with the *Papio* viruses as the host species are in Fig. 3a.

Of the lineages originating directly from the MMF, this leaves only the hominid virus lineage B3 and the PbadLCV-1 line (B4) requiring transfer between host species to account for their placing. The Clade B local timescale of Fig. 4b dates the MMF node to about 12 millions of years before present (MYA), and we thus interpret these features to indicate that the hominid LCV lineage B3 arose from a monkey LCV lineage within the last 12 million years. Because the branching order of the hominid LCVs in B3 does not match that of their hosts (Fig. 3a), development of B3 must have involved minimally two transfer events: one from a monkey to a hominid host, and the other either also from a monkey to a hominid host or between distinct hominid lineages. The precise origin of the B3 lineage is obscured by the MMF, but in principle the line could have arisen from the single lineage immediately ancestral to other, unresolved branchings covered by the MMF, or from within unresolved branchings of the MMF, or from an early point in a monkey LCV lineage descendant from the MMF (not necessarily one of those visible as B4 to B7). The first of these scenarios is weakly supported in the tree of Fig. 4a. Turning to subclade B6 and the three non-macaque viruses therein (CnegLCV-1, PpygLCV-2 and

SsynLCV-2), these all lie in one tightly delimited clade with MfasLCV-1. Notably, MfasLCV-1 and the two hominoid viruses all have hosts from South East Asia, but the host of CnegLCV-1 is African. The multifurcated node giving rise to these viruses has an estimated date of 1.2 MYA.

Discussion

Two experimental lines were followed in the present study to improve our understanding of LCV evolution. First, we searched for additional viruses belonging to Genogroup 2. In this study, >500 specimens were collected from live or deceased individuals of 49 primate species (hominoids, Old World monkeys, New World monkeys, prosimians) from three continents. In a previous survey (Ehlers et al., 2003), >600 specimens were analysed. From both studies, covering a total of >1100 specimens, sequences were detected from 47 distinct novel LCVs. Among these there were no LCVs of chimpanzees or Old World monkeys in Clade C. We have also searched for possible human LCVs in Clade C but found only known variants of EBV, belonging to Clade B (data not shown). Based on these combined data, we conclude that the existence of two distinct and complete Old World primate LCV lineages as previously proposed (Ehlers et al., 2003) is less likely. In a second step, we extended the amounts of sequence information for LCVs by up to 7-fold. This was successful for 25 viruses but failed for the remainder, most likely because of genome copy numbers being too low for LD-PCR. With this much larger data set we re-examined LCV phylogeny, extending and refining the phylogenetic analysis published earlier (Ehlers et al., 2003). In particular, the LCV phylogeny could now be compared with host phylogeny with adequate precision, and the topology of Clade B could be analysed with higher resolution.

Interpretation of the *Lymphocryptovirus* tree as reflecting long term synchronous development with primate host lineages accounts well for the pattern of major branches, and is in harmony with features in other parts of the *Herpesviridae* tree (McGeoch et al., 2006, 2008). However, to account for aspects of relative branch lengths, it is then necessary to propose differing rates of evolutionary change for different regions within the *Lymphocryptovirus* tree. In this connection, we note that phylogenetic analyses typically show a markedly lower rate of change in the whole gamma-1 lineage than in the gamma-2 (for instance, see branch lengths in fig. 2a; also, McGeoch et al., 2006, 2008). We modelled the rate difference proposed between Clade B and other parts of the *Lymphocryptovirus* tree by applying a single decreased rate from the earliest node in Clade B. We regard this as a simple and justified device, while emphasizing that its *ad hoc* nature urges caution in applying the resulting Clade B timeframe. While we have no data on underlying causes of decreased rates of change, both in gamma-1 lineages relative to gamma-2 and in Clade B relative to other parts of gamma-1, it is interesting to speculate: perhaps these phenomena reflect stages in elaboration of the mode of virus existence and latency exemplified by EBV (Young & Rickinson, 2004).

In summary, elements of both cospeciation and horizontal transmission were observed in LCV evolution with the present study. From the perspective of human virology, it is of interest that EBV belongs to a lineage that arose by interspecies transfer from a line of Old World monkey viruses. However, transfer events from monkey to hominid host lineages appear to belong to an era that lies far back in the evolutionary development of humans and great apes: EBV could not be regarded as an agent that is novel to its present day host species. On our interpretation, Clade C (Genogroup 2) contains hominoid virus lines that have developed in long term synchrony with evolution of their host species. Surprisingly, we have not detected any chimpanzee viruses in this group, and neither is any human virus known. No obvious reason is apparent for these absences, although extinction might be a possible scenario. In this context, it should be noted that the two known types of EBV are closely related strains that both belong in Clade B (McGeoch & Gatherer, 2007).

The evolutionary history of EBV, and the observation that EBV can also be experimentally transmitted to foreign primate hosts (Frank et al., 1976; Cleary et al., 1985), indicates that the species specificity of LCVs is not absolute. Rather, LCV, may be horizontally and zoonotically transmittable. The most recent transmissions evidently occurred about 1 MYA from macaques to hominoids (orang utans and gibbons) in Indonesia, resulting in the emergence of PpygLCV2 and SsynLCV2. In the wider picture of the mammalian herpesviruses, horizontal transmissions (either observed in vivo or deduced from evolutionary studies) are not a rare phenomenon. A well known example is the herpes B virus: in its natural host (macaques) it is benign but upon transmission to humans it is associated with high mortality (Palmer, 1987). Also the human Herpes simplex virus can be transferred to nonhuman primates thereby occasionally killing complete groups of captive individuals (Mätz-Rensing et al., 2003). Pseudorabies virus which naturally infects pigs can transmit to dogs, cats and other carnivores with rabies-like symptoms (Mettenleiter 2008). The gamma-2 herpesviruses Alcelaphine herpesvirus 1 and Ovine herpesvirus 2 are both asymptomatic in their hosts (wildebeest and sheep, respectively) but cause malignant catarrhal fever (MCF), an often fatal disease, in cattle (Ackermann, 2006). Ovine herpesvirus 2 also causes an MCF-like disease in pigs (Albini et al., 2003). Several other indications for herpesvirus transmission to foreign hosts have been published (Leendertz et al., 2009; Ehlers et al., 2008; Richman et al., 1999; Huang et al., 1978; Melendez et al., 1969). Taken together, this knowledge indicates a zoonotic potential for herpesviruses, and our present results show that this is also the case for LCVs. Herpesviruses appear not to break the species barrier as often and readily as some RNA viruses. However, as exemplified by the emergence of human immunodeficiency viruses from origins in various simian immunodeficiency viruses (Hahn, 2000), in countries with populations of nonhuman primates, the frequent handling of primates, their meat and organs might facilitate the zoonotic transmission of herpesviruses from monkeys and apes to humans.

Methods

Sample collection and processing, PCR methods and sequence analysis

Over six years, blood, tissue and faeces samples (n=502) were collected from live or deceased individuals of 49 primate species (apes, Old World monkeys, New World monkeys and prosimians) in the Taï National Park of Côte d´Ivoire (Leendertz et al., 2006), Cameroon, Republic of Congo, Democratic Republic of Congo, Uganda, Indonesia and Vietnam. Samples were also collected from live or deceased individuals in several German zoological gardens and primate facilities. The primate species that yielded herpesvirus sequence data are listed in Table 1, and details of the samples are available on request. DNA was prepared with the QiaAmp tissue kit according to the manufacturer's instructions.

For universal detection of herpesviruses, pan-herpes DPOL PCR for amplification of 160-181 bp (excluding primer-binding sites) of the DPOL gene was carried out as described previously (Chmielewicz et al., 2003). For detection of the MDBP gene and the gB gene, LCV sequences were amplified with five degenerate, deoxyinosine-substituted primer sets in a nested format (Supplementary Table 1). These sets were based on the published CalHV-3 BALF2 gene (Rivailler et al., 2002) (primer set BALF2A) and the PtroLCV-1 gB gene (set BALF4A), or on the PpygLCV-1 BALF2 gene (set BALF2B), the PpygLCV-1 BALF4 gene (set BALF4B) and the GgorLCV-2 BALF4 gene (set BALF4C), as determined in this study. The primer binding sites were placed in regions conserved among the gammaherpesviruses. The primers were only minimally degenerate in order to avoid amplification of gamma-2 viruses. PCR was carried out at an annealing temperature of 46^{o}C under conditions used in pan-herpes DPOL PCR. LD-PCR was performed in a nested format with the TaKaRa-Ex PCR system (Takara Bio Inc.) according to the manufacturer's instructions, using virus-specific primers (not listed). PCR products were purified by using the PCR purification kit (QIAgen) and directly sequenced with the Big Dye terminator cycle sequencing kit in a 377 DNA automated sequencer (Applied Biosystems).

Provisional nomenclature, abbreviations and nucleotide sequence accession numbers of novel herpesviruses

Names and abbreviations for newly detected LCVs were formed from the host species name and the genus to which the virus was tentatively assigned (example: **P**an **tro**glodytes **l**ympho**c**ryptO**v**irus, PtroLCV), and are listed with GenBank accession numbers in Table 2. LCVs with published sequences that were used in the analyses and LCVs (Ehlers et al., 2003) from which additional sequences were generated for this study are also listed (Table 2).

Phylogenetic analysis

Amino acid sequence alignments for sets of herpesvirus sequences were made using MAFFT (Katoh et al., 2002). Regions in alignments that were considered too variable to be confidently alignable, plus locations containing a gapping character in any sequence, were removed before using alignments for phylogenetic inference. The amino acid substitution table of Jones et al. (1992) was used in tree inference programs.

Phylogenetic trees were inferred from alignments of amino acid sequences. Preliminary trees were derived by the neighbour-joining method using PROTDIST and NEIGHBOR (PHYLIP suite v 3.63; Felsenstein, 1993). Phylogenetic relationships were investigated in depth by compute-intensive Bayesian analysis with Monte Carlo Markov chains (BMCMC) (MrBayes v 3.1; Ronquist & Huelsenbeck, 2003). Default values for priors were used. For concatenated alignments of gB and DPOL sequences, the two datasets were treated in separate partitions. Substitution rates were modelled as a discrete gamma distribution of four classes plus one invariant class. MrBayes runs were for at least one million generations, and comprised two processes each of one unheated and three heated chains. Trees were sampled every 100 generations and a large burn-in (usually 5001 trees) was applied. Majority rule consensus trees were obtained from the output.

Molecular clock versions of previously derived trees were computed by maximum likelihood methods with CODEML (PAML suite v 4; Yang, 2007), with substitution rates modelled as a discrete gamma distribution of five classes. Timescales for molecular clock gammaherpesvirus trees were applied with a single calibration point proposed by reference to the phylogenetic tree of primate host lineages. The comprehensive study of Bininda-Emonds et al. (2007) was used as the reference for primate phylogeny and divergence dates. It should be noted that divergence dates from Bininda-Emonds et al. are generally older than those from earlier studies (Schneider, 2000; Raaum et al., 2005; Steiper & Young, 2006) which were used in our recent papers (McGeoch et al., 2006, 2008; Ehlers et al., 2008; Leendertz et al., 2009).

Sequence alignments employed in this work are available, with inferred trees, on request to d.gatherer@mrcvu.gla.ac.uk.

Acknowledgements

The authors thank Sonja Liebmann, Nezlisah Yasmum, Claudia Hedemann and Julia Tesch for excellent technical assistance. The supply of primate samples by Kerstin Mätz-Rensing, Andreas Ochs and Benoit deThoisy is kindly acknowledged. For work in Côte d'Ivoire we thank the Ivorian authorities for long-term support, especially the Ministry of the Environment and Forests as well as the Ministry of Research, the directorship of the Taï National Park, and the Swiss Research Centre in Abidjan. For work in Cameroon we thank the Ministry of Forests and Fauna and the staff of the

377 Limbe Wildlife Centre, especially Dr Kiyang John Anyam. For samples from Uganda, we thank the
378 Chimpanzee Sanctuary & Wildlife Conservation Trust for permission to collect samples.

Literature

1. **Ackermann, M. (2006).** Pathogenesis of gammaherpesvirus infections. *Vet Microbiol* **113**, 211-22.
2. **Albini, S., Zimmermann, W., Neff, F., Ehlers, B., Häni, H., Li, H., Hüssy, D., Engels, M., & Ackermann, M. (2003).** Identification and quantification of ovine gammaherpesvirus 2 DNA in fresh and stored tissues of pigs with symptoms of porcine malignant catarrhal fever. *J Clin Microbiol* **41**, 900-4.
3. **Baer, R., Bankier, A. T., Biggin, M. D., Deininger, P. l., Farrell, P. J., Gibson, T. G., Hatfull, G., Hudson, G. S., Satchwell, S. C., Seguin, C., Tuffnell, P. S., & Barrell, B.G. (1984).** DNA sequence and expression of the B95-8 Epstein–Barr virus genome. *Nature* **310**, 207–211.
4. **Bininda-Emonds, O. R. P., Cardillo, M., Jones, K. E., MacPhee, R. D. E., Beck, R. M. D., Grenyer, R., Price, S. A., Vos, R. A., Gittleman, J. L. & Purvis, A. (2007).** The delayed rise of present-day mammals. *Nature* **446**, 507-512.
5. **Cho, Y., Ramer, J., Rivailler, P., Quink, C., Garber, R. L., Beier, D. R., & Wang, F. (2001).** An Epstein-Barr-related herpesvirus from marmoset lymphomas. *Proc Natl Acad Sci U S A* **98**, 1224-1229.
6. **Chmielewicz, B., Goltz, M. Lahrmann, K.H. & Ehlers, B. (2003).** Approaching virus safety in xenotransplantation: A search for unrecognized herpesviruses in pigs. *Xenotransplantation* **10**, 349-357.
7. **Cleary, M. L., Epstein, M. A., Finerty, S., Dorfman, R. F., Bornkamm, G. W., Kirkwood, J. K., Morgan, A. J., & Sklar, J. (1985).** Individual tumors of multifocal EB virus-induced malignant lymphomas in tamarins arise from different B-cell clones. *Science* **228**, 722-724.
8. **Davison, A. J., Eberle, R., Ehlers, B., Hayward, G. S., McGeoch, D. J., Minson, A. C., Pellett, P. E., Roizman, B., Studdert, M. J. & Thiry, E. (2009).** The order *Herpesvirales*. *Arch Virol* **154**, 171-177.
9. **Ehlers, B., Ochs, A., Leendertz, F., Goltz, M., Boesch, C. & Mätz-Rensing, K. (2003).** Novel simian homologues of Epstein-Barr virus. *J Virol* **77**, 10695-10699.
10. **Ehlers, B., Dural, G., Yasmum, N., Lembo, T., de Thoisy, B., Ryser-Degiorgis, M. P., Ulrich, R. G. & McGeoch, D. J. (2008).** Novel mammalian herpesviruses and lineages within the Gammaherpesvirinae: cospeciation and interspecies transfer. *J Virol* **82**, 3509-16.
11. **Epstein, M. A., Achong, B. G. & Barr, Y. M. (1964).** Virus particles in cultured lymphoblasts from Burkitt´s lymphoma. *Lancet i*, 702-703.
12. **Felsenstein, J. (1993).** PHYLIP (Phylogeny Inference Package) version 3.6a2. *Distributed by*

413 *the author. Department of Genetics, University of Washington, Seattle.*
414 13. **Frank, A., Andiman, W. A. & Miller, G. (1976).** Epstein-Barr virus and nonhuman primates:
415 natural and experimental infection. *Adv Cancer Res* **23**, 171-201.
416 14. **Fujimoto, K., Terato, K., Miyamoto, J., Ishiko, H., Fujisaki, M., Cho, F. & Honjo, S. (1990).**
417 Establishment of a B-lymphoblastoid cell line infected with Epstein-Barr-related virus from a
418 cynomolgus monkey (*Macaca fascicularis*). *J Med Primatol* **19**, 21-30.
419 15. **Gerner, C. S., Dolan, A. & McGeoch, D. J. (2004).** Phylogenetic relationships in the
420 *Lymphocryptovirus* genus of the Gammaherpesvirinae. *Virus Res* **99**, 187-192.
421 16. **Hahn, B. H., Shaw, G. M., De Cock, K. M. & Sharp, P. M. (2000).** AIDS as a zoonosis:
422 scientific and public health implications *Science* **287**, 607-14.
423 17. **Hayashi, K., Chen, H. L., Yanai, H., Koirala, T. R., Ohara, N., Teramoto, N., Oka, T.,**
424 **Yoshino, T, Takahashi, K., Miyamotor, K., Fujimoto, K., Yoshikawa, Y. & Akagi, T. (1999).**
425 Cyno-EBV (EBV-related herpesvirus from cynomolgus macaques) induces rabbit malignant
426 lymphomas and their tumor cell lines frequently show specific chromosomal abnormalities. *Lab*
427 *Invest* **79**, 823-835**.**
428 18. **Huang, E. S., Kilpatrick, B., Lakeman, A., & Alford, C. A. (1978).** Genetic analysis of a
429 cytomegalovirus-like agent isolated from human brain. *J Virol* **26**, 718-723.
430 19. **Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992).** The rapid generation of mutation data
431 matrices from protein sequences. *Comp Appl Biosci* **8**, 275-282.
432 20. **Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002).** MAFFT: a novel method for rapid
433 multiple sequence alignment based on fast Fourier transform. *Nucl Acids Res* **30**, 3059-3066.
434 21. **Landon, J. C., L. B. Ellis, V. H. Zeve, and D. P. Fabrizio (1980).** Herpes-type virus in cultured
435 leukocytes from chimpanzees. *J Natl Cancer Inst* **40**, 181-192.
436 22. **Leendertz, F. H., Pauli, G., Ellerbrok, H., Maetz-Rensing, K., Boardman, W., Jensen, S. A.,**
437 **Junglen, S., & Boesch, C. (2006).** Pathogens as drivers of population declines: The
438 importance of systematic monitoring in great apes and other threatened mammals. *Biol*
439 *Conserv* **131**, 325-337.
440 23. **Leendertz, F. H., Deckers, M., Schempp, W., Lankester, F., Boesch, C., Mugisha, L.,**
441 **Dolan, A., Gatherer, D., McGeoch, D. J. & Ehlers, B. (2009).** Novel cytomegaloviruses in
442 free-ranging and captive great apes: phylogenetic evidence for bidirectional horizontal
443 transmission. *J Gen Virol* **90**, 2386-2394.
444 24. **Mätz-Rensing, K., Jentsch, K.D., Rensing, S., Langenhuyzen, S., Verschoor, E., Niphuis.**
445 **H. & Kaup, F. J. (2003).** Fatal Herpes simplex infection in a group of common marmosets
446 (Callithrix jacchus). *Vet Pathol* **40**, 405–411.

25. **McGeoch, D. J., Rixon, F. J. & Davison, A. J. (2006).** Topics in herpesvirus genomics and evolution. *Virus Res* **117**, 90-104.

26. **McGeoch, D. J. & Gatherer, D. (2007).** Lineage structures in the genome sequences of three Epstein-Barr virus strains. *Virology* **359**, 1-5.

27. **McGeoch, D. J., Davison, A. J., Dolan, A., Gatherer, D. & Sevilla-Reyes, E. E. (2008).** Molecular evolution of the *Herpesvirales*. In *Origin and evolution of viruses, 2nd edition*, pp. 447-475. Edited by E. Domingo, C. R. Parrish & J. J. Holland. London: Academic Press.

28. **Mettenleiter, T. C. (2008).** Pseudorabies virus. In: *Encyclopedia of Virology, 3rd edition*, pp. 341-351. Edited by Brian Mahy and Marc Van Regenmortel. Amsterdam: Elsevier.

29. **Meléndez, L. V., Hunt, R. D., Daniel, M. D., García, F. G. & Fraser, C. E. (1969).** Herpesvirus saimiri. II. Experimentally induced malignant lymphoma in primates. *Lab Anim Care* **19**, 378-86.

30. **Neubauer, R. H., Rabin, H., Strand, B. C., Nonoyama, M. & Nelson-Rees, W. A. (1979).** Establishment of a lymphoblastoid cell line and isolation of an Epstein-Barr-related virus of gorilla origin. *J Virol* **31**, 845-848.

31. **Pagano, J. S. (1999).** Epstein-Barr virus: the first human tumor virus and its role in cancer. *Proc Assoc Am Physicians* **111**, 573-580.

32. **Palmer, A. E. (1987).** B virus, Herpesvirus simiae: historical perspective. *J Med Primatol* **16**, 99-130.

33. **Prepens, S., Kreuzer, K.-A., Leendertz, F., Nitsche, A. & Ehlers, B. (2007).** Discovery of herpesviruses in multi-infected primates using locked nucleic acids (LNA) and a bigenic PCR approach. *Virology Journal* **4**, 84.

34. **Ramer, J. C., Garber, R. L., Steele, K. E., Boyson, J. F., O'Rourke, C. & Thomson, J. A. (2000).** Fatal lymphoproliferative disease associated with a novel gammaherpesvirus in a captive population of common marmosets. *Comp Med* **50**, 59-68.

35. **Rangan, S.R., Martin, L. N., Bozelka, B. E., Wang, N. & Gormus, B. J. (1986).** Epstein-Barr virus-related herpesvirus from a rhesus monkey (*Macaca mulatta*) with malignant lymphoma. *Int J Cancer* **38**, 425-432.

36. **Rasheed, S., Rongey, R. W., Bruszweski, J., Nelson-Rees, W. A., Rabin, H., Neubauer, R. H., Esra, G. & Gardner, M. B. (1977).** Establishment of a cell line with associated Epstein-Barr-like virus from a leukemic orangutan. *Science* **198**, 407-409.

37. **Raaum, R. L., Sterner, K. N., Noviello, C. M., Stewart, C.-B. & Disotell, T. R. (2005).** Catarrhine primate divergence dates estimated from complete mitochondrial genomes: concordance with fossil and nuclear DNA evidence. *J Hum Evol* **48**, 237–257.

480 38. **Richman, L. K., Montali, R. J., Garber, R. L., Kennedy, M. A., Lehnhardt, J., Hildebrandt,**
481 **T., Schmitt, D., Hardy, D., Alcendor, D. J. & Hayward, G. S. (1999).** Novel endotheliotropic
482 herpesviruses fatal for Asian and African elephants. *Science* **283**, 1171-1176.

483 39. **Rivadeneira, E. D., Ferrari, M. G., Jarrett, R. F., Armstrong, A. A., Markham, P., Birkebak,**
484 **T., Takemoto, S., Johnson-Delaney, C., Pecon-Slattery, J., Clark, & Franchini, G. (1999).**
485 A novel Epstein-Barr-Virus-like virus, HV_{MNE}, in a *Macaca nemestrina* with mycosis fungoides.
486 *Blood* **15**, 2090-2101.

487 40. **Rivailler P., Cho Y. G. & Wang F. (2002).** Complete genomic sequence of an Epstein-Barr
488 virus-related herpesvirus naturally infecting a new world primate: a defining point in the
489 evolution of oncogenic lymphocryptoviruses. *J Virol* **76**, 12055-12068.

490 41. **Ronquist, F. & Huelsenbeck, J. P. (2003).** MrBayes 3: Bayesian inference under mixed
491 methods. *Bioinformatics* **19**, 1572-1574.

492 42. **Schneider, H. (2000).** The current status of the new world monkey phylogeny. *Ann Acad Bras*
493 *Ci* **72**, 165-172.

494 43. **Steiper, M. E. & Young, N. M. (2006).** Primate molecular divergence dates. *Mol Phylogenet*
495 *Evol* **41**, 384-394.

496 44. **Vasiljeva, V. A., Markarjan, D. S., Lapin, B. A., Yakovleva, L. A., Ivanov, M. T.,**
497 **Schekolodkin, V. F. & Dzikidze, E. K. (1974).** Establishment of continuous cell lines from
498 leukocytes culture of a hamadryas baboon with leukosis-reticulosis. *Neoplasma* **21**, 537-544.

499 45. **Yang, Z. (2007).** PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**,
500 1586-1591.

501 46. **Young, L. S. & Rickinson, A. B. (2004).** Epstein-Barr virus: 40 years on. *Nat Rev Cancer* **4**,
502 757-768.

Table 1 Primate hosts of lymphocryptoviruses

Host Family, Subfamily, Genus	Host species	
Catarrhini (Old world monkeys and apes)		
Family *Hominidae*		
Genus *Gorilla*	Gorilla gorilla	Gorilla
Genus *Pan*	Pan troglodytes	Chimpanzee
	Pan paniscus	Bonobo
Genus *Pongo*	Pongo pygmaeus	Borneo orangutan
Family *Hylobatidae*		
Genus *Hylobates*	Hylobates lar	White-handed gibbon
	Hylobates muelleri	Mueller´s gibbon
Genus *Symphalangus*	Symphalangus syndactylus	Siamang
Family *Cercopithecidae*		
Subfamily *Cercopithecinae*		
Genus *Cercocebus*	Cercocebus atys	Sooty mangabey
Genus *Cercopithecus*	Cercopithecus hamlyni	Owl-faced monkey
	Cercopithecus cephus	Moustached guenon
	Cercopithecus neglectus	De Brazza´s monkey
	Cercopithecus nictitans	Greater spot-nosed guenon
Genus *Chlorocebus*	Chlorocebus aethiops	Vervet monkey
Genus *Erythrocebus*	Erythrocebus patas	Patas monkey
Genus *Lophocebus*	Lophocebus albigena	Grey-cheeked mangabey
	Lophocebus aterrimus	Black mangabey
Genus *Macaca*	Macaca fascicularis	Long-tailed macaque
	Macaca thibetana	Tibetan stump-tailed macaque
	Macaca fuscata	Japanese macaque
	Macaca mulatta	Rhesus monkey
Genus *Mandrillus*	Mandrillus sphinx	Mandrill
Genus *Miopithecus*	Miopithecus talapoin	Dwarf guenon
Genus *Papio*	Papio anubis	Olive baboon
	Papio hamadryas	Hamadryas baboon
Subfamily *Colobinae*		
Genus *Colobus*	Colobus guereza	Black-and-white colobus
	Colobus polykomos	King colobus
Genus *Piliocolobus*	Piliocolobus badius	Red colobus
Genus *Semnopithecus*	Semnopithecus entellus	Hanuman langur
Platyrrhini (New world monkeys)		
Family *Atelidae*	Ateles paniscus	Black spider monkey
Family Cebidae	Callithrix penicillata	Black-pencilled marmoset
	Callithrix jacchus	Common marmoset
	Leontopithecus rosalia	Golden lion tamarin
	Saimiri sciureus	Common squirrel monkey
Family *Pitheciidae*	Pithecia pithecia	White-faced saki

1 **Table 2 Viruses, abbreviations and GenBank accession numbers**

Host species	Origin of host	Virus name	Abbreviation	Novel virus[#]	gB-DPOL Sequence[§]	Accession number
Viruses from this study						
Old World primates						
Sooty mangabey	Ivory coast	Cercocebus atys lymphocryptovirus 1	CatyLCV-1	x		GQ921921
Owl-faced monkey	Germany (Zool. Gardens)	Cercopithecus hamlyni lymphocryptovirus 1	ChamLCV-1	x		AY608706
Moustached guenon	Cameroon	Cercopithecus cephus lymphocryptovirus 1	CcepLCV-1	x		AY608711
De Brazza´s monkey	Cameroon	Cercopithecus neglectus lymphocryptovirus 1	CnegLCV-1	x	x	AY728176
De Brazza´s monkey	Cameroon	Cercopithecus neglectus lymphocryptovirus 2	CnegLCV-2	x		AY608712
Greater spot-nosed guenon	Cameroon	Cercopithecus nictitans lymphocryptovirus 1	CnicLCV-1	x		AY608709
Vervet monkey	Germany (Primate facility)	Chlorocebus aethiops lymphocryptovirus 1	CaetLCV-1	x		AY608702
Vervet monkey	Germany	Chlorocebus aethiops lymphocryptovirus 2	CaetLCV-2	x		GQ921922
Black-and-white colobus	Cameroon; Germany (Zool. Gardens; Primate facility)	Colobus guereza lymphocryptovirus 1	CgueLCV-1		x	AF534219
King colobus	Ivory coast	Colobus polykomos lymphocryptovirus 1	CpolLCV-1	x		GQ921923
Patas monkey	Cameroon; Germany (Primate facility)	Erythrocebus patas lymphocryptovirus 1	EpatLCV-1		x	AY196148
Gorilla	Congo; Cameroon; Germany, Belgium, USA (Zool. Gardens); Cell line	Gorilla gorilla lymphocryptovirus 1	GgorLCV-1		x	AF534225
Gorilla	Congo; Cameroon; Germany (Zool. Gardens)	Gorilla gorilla lymphocryptovirus 2	GgorLCV-2		x	AY129395
White-handed gibbon	Germany (Zool. Gardens)	Hylobates lar lymphocryptovirus 1	HlarLCV-1		x	AY196147
Mueller´s gibbon	Germany (Zool. Gardens)	Hylobates muelleri lymphocryptovirus 1	HmueLCV-1	x		AY273184
Grey-cheeked mangabey	Cameroon	Lophocebus albigena lymphocryptovirus 1	LalbLCV-1	x		AY608710
Black mangabey	Germany (private husbandry)	Lophocebus aterrimus lymphocryptovirus 1	LateLCV-1		x	AY174067
Long-tailed macaque	Germany (Primate facility, Zool. Gardens)	Macaca fascicularis lymphocryptovirus 1	MfasLCV-1		x	AF534221
Japanese macaque	Germany (Primate facility)	Macaca fuscata lymphocryptovirus 1	MfusLCV-1		x	AF534224
Japanese macaque	Germany (Primate facility)	Macaca fuscata lymphocryptovirus 2	MfusLCV-2		x	AY172954
Tibetan stump-tailed macaque	Germany (Primate facility)	Macaca tibetana lymphocryptovirus 2	MtibLCV-2	x	x	GQ921925
Mandrill	Germany (Primate facility, Zool. Gardens)	Mandrillus sphinx lymphocryptovirus 1	MsphLCV-1		x	AF534227
Mandrill	Cameroon; Germany (Zool. Garden)	Mandrillus sphinx lymphocryptovirus 2	MsphLCV-2	x		AY728172
Dwarf Guenon	Cameroon	Miopithecus talapoin lymphocryptovirus 1	MtalLCV-1	x		AY608708

2

……..Table 2 continued

Host species	Origin of host	Virus name	Abbreviation	Novel virus	gB+DPOL Seq.	Accession number
Bonobo	South Africa; Germany (Zool. Gardens)	Pan paniscus lymphocryptovirus 1	PpanLCV-1		x	AF534220
Chimpanzee	Ivory Coast; Uganda; Congo; South Africa; Germany (Primate facility; Zool. Garden); Cell line	Pan troglodytes lymphocryptovirus 1	PtroLCV-1		x	AF534226
Olive baboon	Cameroon; Tanzania	Papio anubis lymphocryptovirus 1	PanuLCV-1	x		AY728174
Hamadryas baboon	Tanzania; Germany (Zool. Gardens)	Papio hamadryas lymphocryptovirus 2	PhamLCV-2		x	AF534229
Red colobus	Ivory coast	Piliocolobus badius lymphocryptovirus 1	PbadLCV-1		x	AF534228
Red colobus	Democratic Republic Congo	Piliocolobus badius lymphocryptovirus 2	PbadLCV-2	x		GQ921927
Orang utan	Indonesia; Germany (Zool. Gardens)	Pongo pygmaeus lymphocryptovirus 1	PpygLCV-1		x	AY129398
Orang utan	Germany (Zool. Gardens)	Pongo pygmaeus lymphocryptovirus 2	PpygLCV-2	x	x	GQ921926
Hanuman langur	Germany (Zool. Gardens)	Semnopithecus entellus lymphocryptovirus 1	SentLCV-1		x	AF534223
Siamang	Germany (Primate facility)	Symphalangus syndactylus lymphocryptovirus 1	SsynLCV-1		x	AY608703
Siamang	Germany (Primate facility)	Symphalangus syndactylus lymphocryptovirus 2	SsynLCV-2	x	x	GQ921924
New World primates						
Black spider monkey	Germany (Zool. Gardens)	Ateles paniscus lymphocryptovirus 1	ApanLCV-1		x	AY139028
Black-pencilled marmoset	Germany (Primate facility)	Callithrix penicillata lymphocryptovirus 1	CpenLCV-1		x	AY139026
Golden-Lion tamarin	Germany (Primate facility)	Leontopithecus rosalia lymphocryptovirus 1	LrosLCV-1	x		AY608705
White-faced saki	Germany (Zool. Gardens)	Pithecia pithecia lymphocryptovirus 1	PpitLCV-1		x	AY139025
Common squirrel monkey	French Guinea; Germany (Zool. Gardens)	Saimiri sciureus lymphocryptovirus 2	SsciLCV-2		x	AY139024
Common squirrel monkey	Germany (Zool. Gardens)	Saimiri sciureus lymphocryptovirus 3	SsciLCV-3	x		AY854172
Published viruses						
Complete genome						
Human		Epstein-Barr virus	EBV = HHV-4			NC_007605
Rhesus monkey (Macaca mulatta)		Rhesus monkey lymphocryptovirus	RLV = CeHV-15			AY037858
Marmoset (Callithrix jacchus)		Callitrichine herpesvirus 3	CalHV-3			AF319782
gB - DPOL sequence						
Hamadryas baboon	Tanzania; Germany	Papio hamadryas lymphocryptovirus 3	PhamLCV-3			EU118146

[#] The viruses which were discovered in the course of the present study are marked

[§] The viruses for which we determined contiguous gB plus DPOL seuquences are marked

Figure legends

Figure 1. Map of amplified genes and diagrams of PCR strategies Figure 1.
Degenerate nested primers were used to amplify part of the MDBP gene, the gB gene or the DPOL gene. The amplified fragments are represented by thin solid lines between the primer binding sites (black triangles). Long-distance nested PCR was performed with specific primers. The amplified fragments are represented by dashed lines between the primer binding sites (open triangles). The numbers of gB to DPOL and MDBP to DPOL sequences are specified, and their locations are depicted with thick solid lines. At the top of the figure, the genomic locus spanning ORF BALF2 (MDBP) to ORF BALF5 (DPOL) is depicted with black arrows. The arrowhead indicates the direction of transcription. The ORF designation is adapted from the ORF nomenclature of EBV. The start of the ruler corresponds with the first base of the ORF BALF2.

Figure 2. Phylogenetic trees for the *Herpesviridae* family and the *Lymphocryptovirus* genus.
(a) Overview tree based on a 641 residue alignment of gB amino acid sequences for 58 herpesviruses (listed in Supplementary Table 2). The tree was obtained by the neighbour joining method and is midpoint rooted. The *Alphaherpesvirinae* and *Betaherpesvirinae* subfamilies' branches are shown in summary form, with regions containing multiple branches reduced to single heavy lines. In the *Gammaherpesvirinae*, the Gamma-2 subgroup is also shown in summary form, and for the Gamma-1 subgroup the three major clades (A, B and C) are depicted.
(b) BMCMC tree for the *Lymphocryptovirus* genus. A 946 residue alignment of concatenated partial gB and DPOL amino acid sequences for 28 LCVs plus 11 gamma-2 primate herpesviruses was constructed and analysed by BMCMC. The gamma-2 sequences (listed in Supplementary Table 3) served as an outgroup to locate the root for the LCV clade, and are not shown in the figure. The LCV tree is shown as a majority rule consensus tree. Branch labels for LCVs with New World monkey hosts are in black, for those with Old World monkey hosts are in red, and for those with ape or human hosts are in blue. The major clades (A, B and C) are labelled. There are three multifurcations in Clade B, which represent loci that were not resolved by the BMCMC process. All resolved nodes have posterior probability of 1.00, except for the three marked with filled black circles, which have posterior probabilities in the range 0.71 to 0.78. A scale indicating divergence, as substitutions per site, is at the foot.

Figure 3. Comparison of phylogenetic trees for primate hosts and LCVs.
(a) Tree for primate species that appear as hosts of LCVs in this paper. Data for this tree were extracted from the study of Bininda-Emonds et al. (2007). New World monkey species are listed in black, Old World monkey species in red, and apes plus humans in blue (*C. aterrimus* does not

feature as a host until Fig. 4b). A timescale is shown at the foot, as millions of years before present (MYA).

(b) Molecular clock tree for LCVs. The tree shown was computed from the 39 LCVs tree of Fig. 2b, with imposition of a global molecular clock. The tree in panel b has been scaled to have the same size on the page for divergence of the Old and New World LCVs as that in panel a has for divergence of the Old and New World primates. The tentative timescale (in gray) for panel b is transferred from panel a; it assumes the same date for separation of Old and New World LCV lineages as for separation of Old and New World primate lineages.

Figure 4. Phylogenetic tree for Clade B LCVs.

(a) Best achievable BMCMC tree for Clade B LCVs. A 1042 residue alignment of concatenated partial gB and DPOL amino acid sequences for 28 LCVs was constructed and analysed by BMCMC. Clades A and C were used to provide a root for Clade B, and are not shown. The Clade B tree is shown as a majority-rule consensus tree, with one multifurcated locus. Resolved nodes had posterior probability of 1.00, except for six which had lower posterior probabilities; the posterior probability (as %) is shown to the right of each of these six nodes. LCV names are coloured as for Figs 2 and 3. A scale indicating divergence, as substitutions per site, is at the foot.

(b) Molecular clock tree for Clade B LCVs. The tree shown in panel a was reduced to a multifurcated version at all nodes with posterior probability less than 0.91, and a molecular clock version computed (including Clades A and C). Two additional LCVs (CateLCV-1 and SsynLCV-2, for which lesser amounts of data were available) were interpolated into the molecular clock tree on the basis of additional tree building exercises, and are shown with their terminal branches as dashed lines. The node at the base of the major multifurcated clade (MMF) is marked with a heavy arrow. Subclades are annotated as B1 to B7. The tentative timescale (in gray) sets the node for divergence of the SentLCV-1 lineage from other Clade B LCVs as corresponding to the divergence of the *S. entellus* lineage from lineages leading to species of *Macaca*, *Papio*, etc.

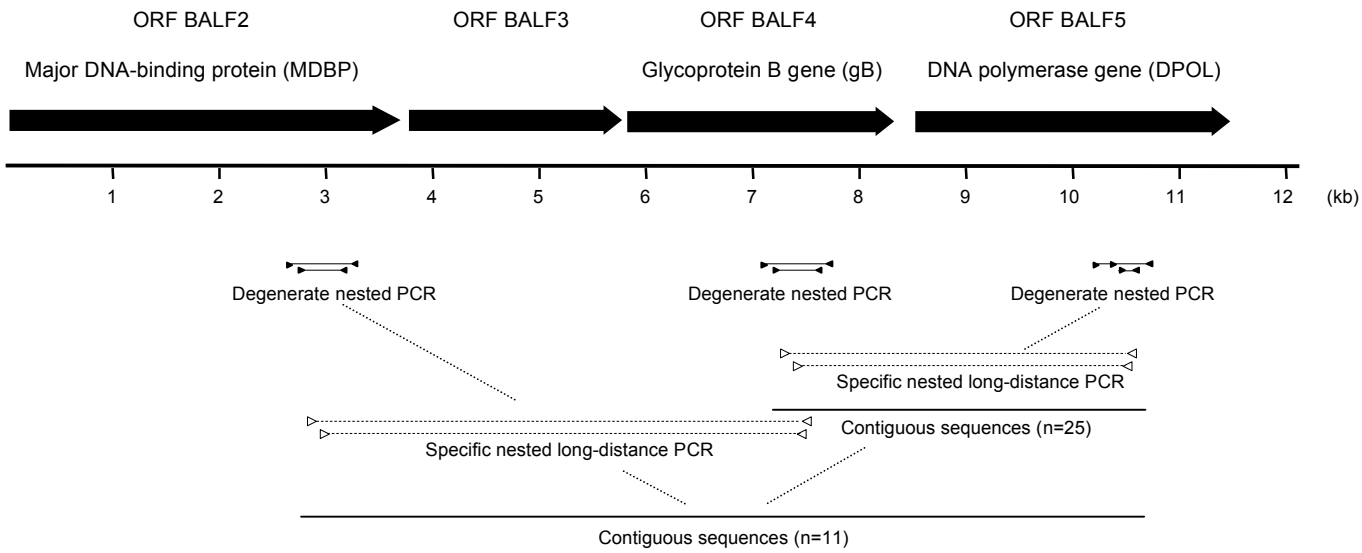
ORF BALF2
ORF BALF3
ORF BALF4
ORF BALF5
Major DNA-binding protein (MDBP)
Glycoprotein B gene (gB)
DNA polymerase gene (DPOL)
1
2
3
4
5
6
7
8
9
10
11
12
(kb)
Degenerate nested PCR
Degenerate nested PCR
Degenerate nested PCR
Specific nested long-distance PCR
Contiguous sequences (n=25)
Specific nested long-distance PCR
Contiguous sequences (n=11)

Figure 1

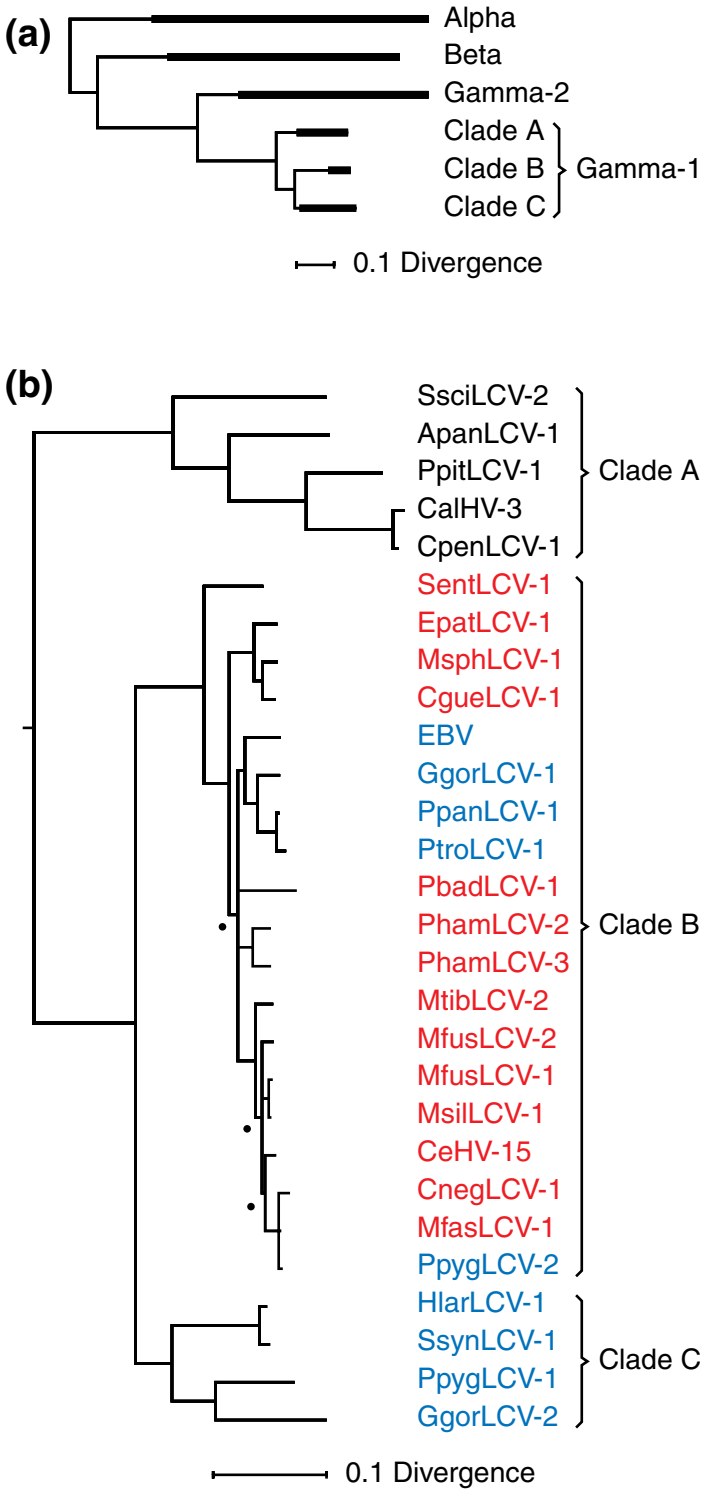
(a)
Alpha
Beta
Gamma-2
Clade A
Clade B
Clade C
Gamma-1
0.1 Divergence
(b)
SsciLCV-2
ApanLCV-1
PpitLCV-1
CalHV-3
CpenLCV-1
Clade A
SentLCV-1
EpatLCV-1
MsphLCV-1
CgueLCV-1
EBV
GgorLCV-1
PpanLCV-1
PtroLCV-1
PbadLCV-1
PhamLCV-2
PhamLCV-3
MtibLCV-2
MfusLCV-2
MfusLCV-1
MsilLCV-1
CeHV-15
CnegLCV-1
MfasLCV-1
PpygLCV-2
Clade B
HlarLCV-1
SsynLCV-1
PpygLCV-1
GgorLCV-2
Clade C
0.1 Divergence

Figure 2

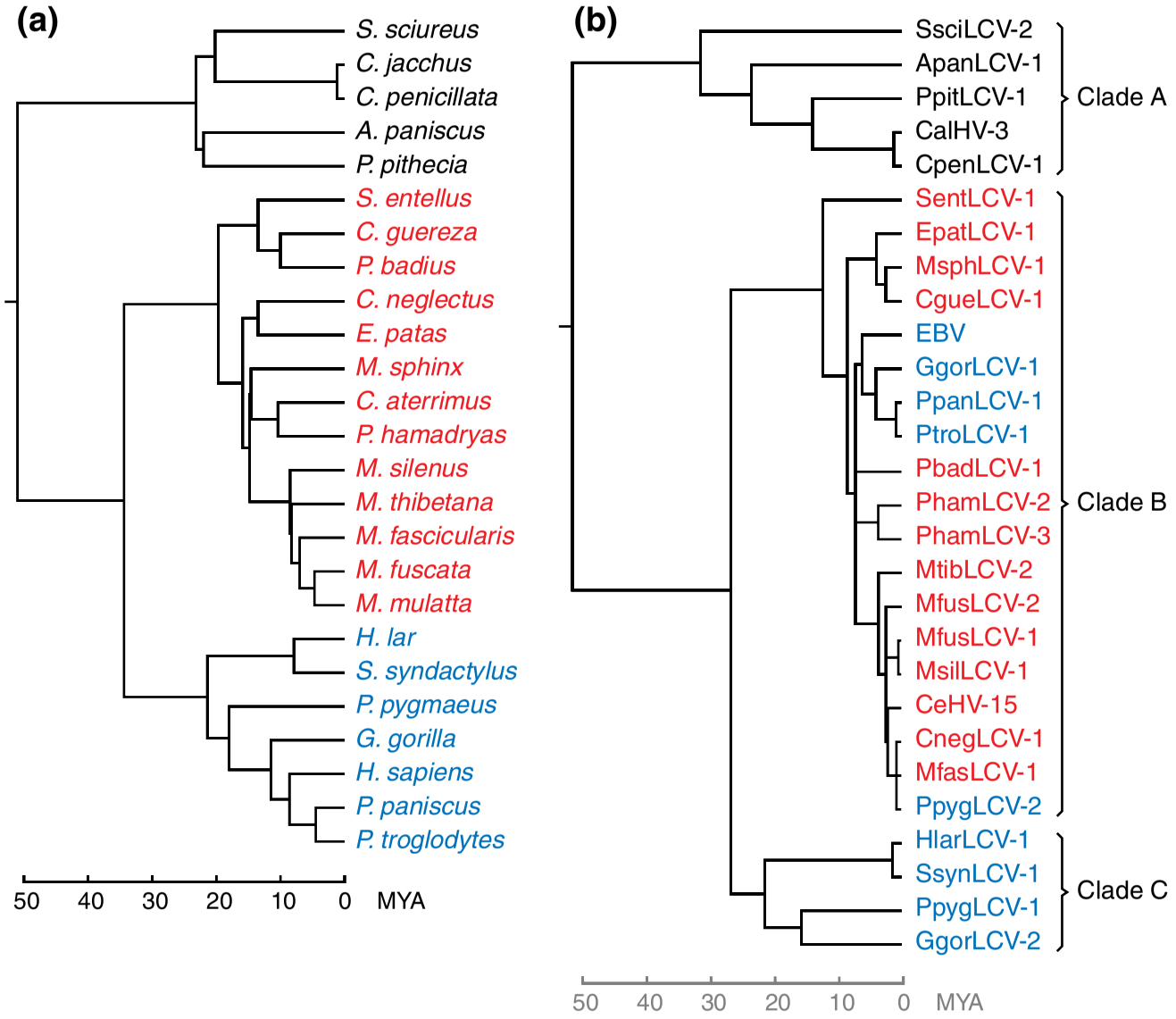
(a)
S. sciureus
C. jacchus
C. penicillata
A. paniscus
P. pithecia
S. entellus
C. guereza
P. badius
C. neglectus
E. patas
M. sphinx
C. aterrimus
P. hamadryas
M. silenus
M. thibetana
M. fascicularis
M. fuscata
M. mulatta
H. lar
S. syndactylus
P. pygmaeus
G. gorilla
H. sapiens
P. paniscus
P. troglodytes
50 40 30 20 10 0 MYA
(b)
SsciLCV-2
ApanLCV-1
PpitLCV-1
CalHV-3
CpenLCV-1
Clade A
SentLCV-1
EpatLCV-1
MsphLCV-1
CgueLCV-1
EBV
GgorLCV-1
PpanLCV-1
PtroLCV-1
PbadLCV-1
PhamLCV-2
PhamLCV-3
MtibLCV-2
MfusLCV-2
MfusLCV-1
MsilLCV-1
CeHV-15
CnegLCV-1
MfasLCV-1
PpygLCV-2
Clade B
HlarLCV-1
SsynLCV-1
PpygLCV-1
GgorLCV-2
Clade C
50 40 30 20 10 0 MYA

Figure 3

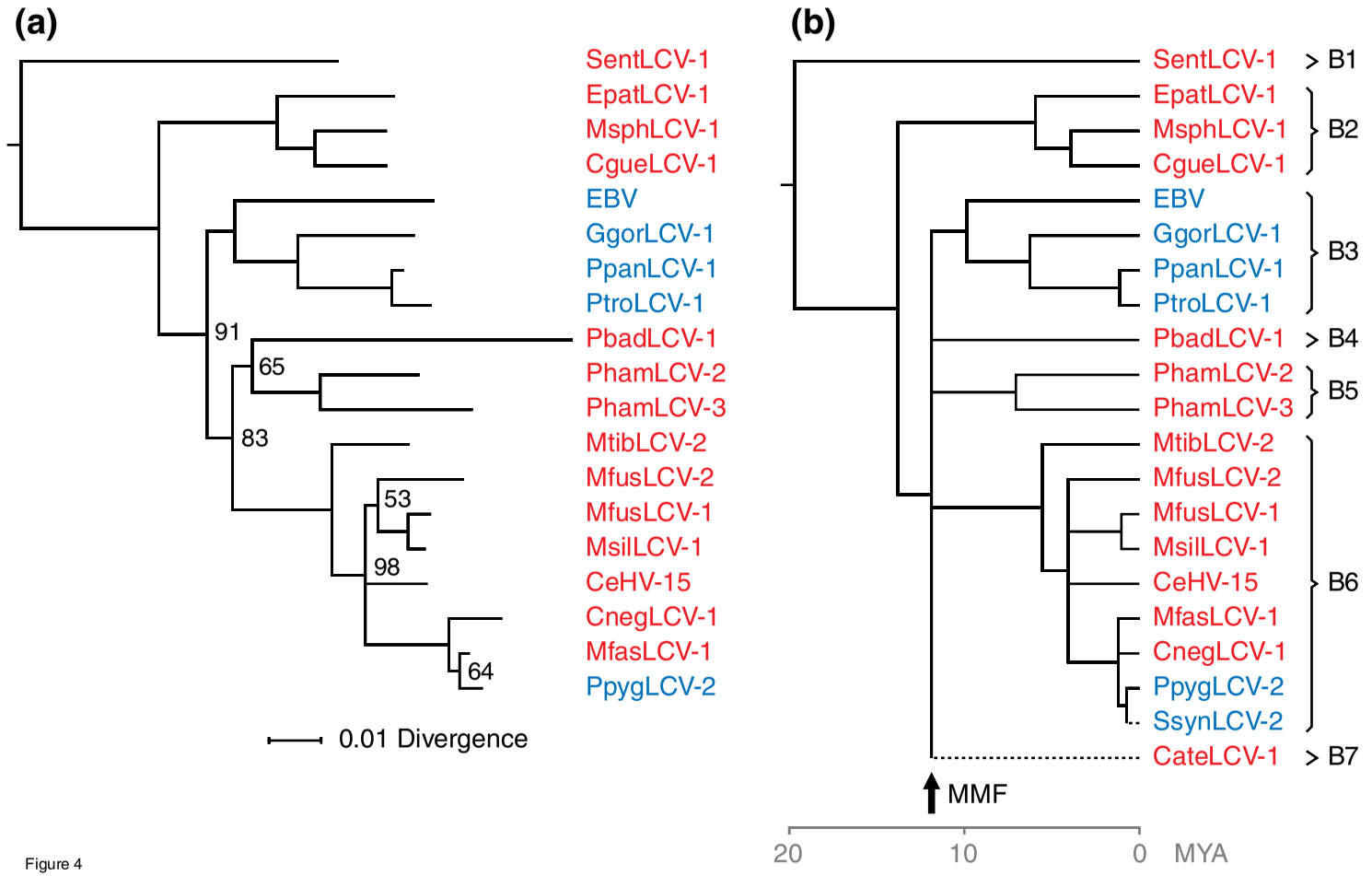
(a)
SentLCV-1
EpatLCV-1
MsphLCV-1
CgueLCV-1
EBV
GgorLCV-1
PpanLCV-1
PtroLCV-1
PbadLCV-1
PhamLCV-2
PhamLCV-3
MtibLCV-2
MfusLCV-2
MfusLCV-1
MsilLCV-1
CeHV-15
CnegLCV-1
MfasLCV-1
PpygLCV-2
91
65
83
53
98
64
0.01 Divergence
(b)
SentLCV-1
EpatLCV-1
MsphLCV-1
CgueLCV-1
EBV
GgorLCV-1
PpanLCV-1
PtroLCV-1
PbadLCV-1
PhamLCV-2
PhamLCV-3
MtibLCV-2
MfusLCV-2
MfusLCV-1
MsilLCV-1
CeHV-15
MfasLCV-1
CnegLCV-1
PpygLCV-2
SsynLCV-2
CateLCV-1
B1
B2
B3
B4
B5
B6
B7
MMF
20
10
0
MYA

Figure 4

Supplementary Table 1

Degenerate primers for amplification of BALF2 and BALF4 sequences

Primer-set abbreviation	Targeted gene	Name of primer [$]	sequence 5´- 3´
BALF2A	BALF2	2944-s	TTCCCGGGGATCTGAT(n/i)AC(n/i)TAYGC
		2944-as	TTTGTCCAGGAAGTAGGTT(n/i)CTRTC(n/i)AC
		2945-s	ACTGTCCCAGCTAGGGA(n/i)TAYCC(n/i)CA
		2945-as	CTGTCTACCCACCTACCCATRAARTA(n/i)CC
BALF2B	BALF2	3055-s	ACAACTACCACAAAGT(n/i)CT(n/i)TTyCC
		3058-as	GACCTGCAGCAGGTTyCTrTC(n/i)AC
		3056-s	TCTGTGGCATAATTTykC(n/i)AGG
		3057-as	CACCCCGGCCCATrAArTA(n/i)CC
BALF4A	BALF4	2753-s	CCATCCAGATCCArTwyGC(n/i)TAyGA
		2756-as	TGGCTGCCAAGCG(n/i)(n/i)T(n/i)GG(n/i)GA
		2754-s	GATGTTCTGCGCCTGRWARTTRTAYTC
		2755-as	GATGTTCTGCGCCTRRWARTTRTA
BALF4B	BALF4	3033-s	CATCGCCAG(S)GC(S)TGGTGC
		3036-as	TCGGCCGAAAC(R)GTCTT(R)A(M)GTG
		3034-s	GGAGCAGCGCAG(N/I)CA(R)AA(Y)ATGGT
		3035-as	CTTCCCCGACAG(R)AAGTAGT(R)CTG
BALF4C	BALF4	3951-s	AGGGCACGGCyAGyTT(y/i)GT
		3951-as	GCCTCTGGTCCwGGCACCA
		3952-s	CGGAGGATTGTTrTTrGC(n/i)TGG
		3952-as	GCGCAACATCTGGTTrAT(y/i)TG

I = Inosine; [$] s=sense, as=antisense

Supplementary Table 2

Viruses and accession numbers for gB tree of 58 viruses (Fig. 2a)

Alphaherpesvirinae	
Bovine herpesvirus 1	NC_001847
Bovine herpesvirus 5	NC_005261
Equine herpesvirus 1	NC_001491
Equine herpesvirus 4	NC_001844
Fibropapilloma-associated turtle herpesvirus	AY644454
Herpes simplex virus 1	NC_001806
Herpes simplex virus 2	NC_001798
Herpesvirus papio 2	NC_007653
Herpesvirus of turkeys	NC_002641
Infectious laryngotracheitis virus	NC_006623
Marek's disease virus 1	NC_002229
Marek's disease virus 2	NC_002577
Psittacine herpesvirus	NC_005264
Pseudorabies virus	NC_006151
Simian agent 8	NC_006560
Simian B virus	NC_004812
Simian varicella virus	NC_002686
Varicella-zoster virus	NC_001348
Betaherpesvirinae	
Aotine cytomegalovirus	FJ483970
Chimpanzee cytomegalovirus	NC_003521
Colburn cytomegalovirus	FJ483969
Guinea pig cytomegalovirus	NC_011587
Human cytomegalovirus	NC_006273
Human herpesvirus 6	NC_001664
Human herpesvirus 7	NC_001716
Murine cytomegalovirus	NC_004065
Rat cytomegalovirus	NC_002512
Rhesus cytomegalovirus	NC_006150
Simian cytomegalovirus	FJ483968
Squirrel monkey cytomegalovirus	FJ483967
Tupaia herpesvirus	NC_002794
Gammaherpesvirinae (gamma-1)	
Callitrichine herpesvirus 3	Genbank acc.-no. in Table 2
Cercopithecine herpesvirus 12	"
Cercopithecine herpesvirus 15	"
Epstein-Barr virus	"
GgorLCV-1	"
GgorLCV-2	"
HlarLCV-1	"
PhamLCV-2	"
PpanLCV-1	"
PpitLCV-1	"
PpygLCV-1	"
PpygLCV-2	"
PtroLCV-1	"
SsciLCV-2	"
SsynLCV-1	"
Gammaherpesvirinae (gamma-2)	
Alcelaphine herpesvirus 1	NC_002531
Bovine herpesvirus 4	NC_002665
Equine herpesvirus 2	NC_001650
Human herpesvirus 8	NC_003409
Herpesvirus ateles	NC_001987
Herpesvirus saimiri	NC_001350
M. fuscata rhadinovirus 1	AY528864
M. mulatta (Rhesus) rhadinovirus 26-95	AF210726
M. mulatta (Rhesus) rhadinovirus 17577	AF083501
Murine herpesvirus 68	NC_001826
Ovine herpesvirus 2	NC_007646
Porcine lymphotropic herpesvirus 1	AF478169

Supplementary Table 3

Gamma-2 viruses and accession numbers used for outgroup to tree of Fig. 2b

G. gorilla rhadinovirus 1	AY177144
Human herpesvirus 8	Genbank acc.-no. in Supplementary Table 2
Herpesvirus ateles	"
Herpesvirus saimiri	"
M. fascicularis rhadinovirus 1	AY138583
M. fascicularis rhadinovirus 2	EU085377
M. fuscata rhadinovirus 1	Genbank acc.-no. in Supplementary Table 2
M. mulatta (Rhesus) rhadinovirus 26-95	"
P. troglodytes rhadinovirus 1	AY138585
P. troglodytes rhadinovirus 2	EU085378
S. sciureus gammaherpesvirus 2	AY138584