THE ROYAL SOCIETY

doi 10.1098/rstb.2000.0775

Molecular evolution of the γ-*Herpesvirinae*

Duncan J. McGeoch

Medical Research Council Virology Unit, Institute of Virology, University of Glasgow, Church Street, Glasgow G11 5JR, UK
(*d.mcgeoch@vir.gla.ac.uk*)

Genomic sequences available for members of the γ-*Herpesvirinae* allow analysis of many aspects of the group's evolution. This paper examines four topics: (i) the phylogeny of the group; (ii) the histories of γ-herpesvirus-specific genes; (iii) genomic variation of human herpesvirus 8 (HHV-8); and (iv) the relationship between Epstein–Barr virus types 1 and 2 (EBV-1 and EBV-2). A phylogenetic tree based on eight conserved genes has been constructed for eight γ-herpesviruses and extended to 14 species with smaller gene sets. This gave a generally robust assignment of evolutionary relationships, with the exception of murine herpesvirus 4 (MHV-4), which could not be placed unambiguously on the tree and which has evidently experienced an unusually high rate of genomic change. The γ-herpesviruses possess a variable complement of genes with cellular homologues. In the clearest cases these virus genes were shown to have originated from host genome lineages in the distant past. HHV-8 possesses at its left genomic terminus a highly diverse gene (*K1*) and at its right terminus a gene (*K15*) having two diverged alleles. It was proposed that the high diversity of *K1* results from a positive selection on *K1* and a hitchhiking effect that reduces diversity elsewhere in the genome. EBV-1 and EBV-2 differ in their alleles of the *EBNA-2*, *EBNA-3A*, *EBNA-3B* and *EBNA-3C* genes. It was suggested that EBV-1 and EBV-2 may recombine in mixed infections so that their sequences outside these genes remain homogeneous. Models for genesis of the types, by recombination between diverged parents or by local divergence from a single lineage, both present difficulties.

Keywords: phylogenetic tree; co-speciation; gene capture; hitchhiking effect; recombination; *Herpesviridae*

1. INTRODUCTION

My aim in this paper is to present and discuss current data and ideas concerning the evolution of the γ-*Herpesvirinae* subfamily of the family *Herpesviridae*. Present understanding is that the γ-herpesviruses developed as a distinct lineage in the distant past, perhaps around 200 million years (Myr) ago (McGeoch *et al.* 1995). Here I treat aspects of the evolutionary development of that lineage, but I do not deal with ideas on earlier evolution of the *Herpesviridae*. The emphasis of my treatment is on molecular evolution of the virus genomes, in terms of their gene sequences and gene contents, an approach made possible by the large amounts of DNA sequences for herpesviruses and their hosts that have accumulated over the last two decades. The picture that emerges is detailed and complex, and also inevitably incomplete. A long-term goal in studying the evolution of virus pathogens must be to understand the intricacies of their history and behaviour in functional terms, but for the present we are limited largely to a descriptive account of likely events in the history without real detail on the deeper 'why?' There are four major sections in this paper. First is an account of γ-herpesvirus phylogeny and next a discussion of genes that within the herpesvirus family are specific to γ-herpesviruses. These are followed by two sections that treat specific aspects of variation in the two known human γ-herpesviruses, EBV and HHV-8 (all abbreviations for virus names are given in tables 1 and 2). The treatment is thus selective, and a number of other topics could have been pursued; a brief listing of these is given at the end of § 6.

As at mid-2000, complete genome sequences are available for eight γ-herpesvirus species, listed in table 1. These provide the main source for evaluating the evolutionary history of the subfamily, together with single gene sequences for other viruses (see table 2) and studies of sequence variation within single virus species. Inspection of the organization and gene contents of the sequenced genomes gives two contrasting impressions: conservation and divergence. The overall genomic layout is conserved, with a long region of unique sequence bounded by terminal repeat elements, and with the presence and genomic order of 42 genes that belong to the ancestral core set of mammalian herpesviruses. However, there are two respects in which the genomes are well diverged from each other. First, sequences of orthologous genes have undergone extensive processes of nucleotide substitution and local addition/deletion change. Second, non-core genes (including γ-specific genes) comprise about half of the total in each genome and the presence or absence of γ-specific genes varies with each genome, so that no two viruses possess an identical complement of this class. The range of genome sizes and base compositions (table 1) also emphasizes the diversity of these viruses. From the viewpoint of genomic comparison, the sets of core genes and of other genes contribute in distinct ways to defining the γ-*Herpesvirinae*: the core genes in terms of their

Table 1. *Completely sequenced γ-herpesvirus genomes*

(Abbreviation: HV, herpesvirus.)

virus species[a]	virus name abbreviation	genome size kbp[b]	base composition (G + C)[b]	references and accession numbers
Epstein–Barr virus (human HV-4)	EBV[c]	154	59	Baer *et al.* (1984); V01555 Parker *et al.* (1990); M35547
human HV-8 (Kaposi's sarcoma-associated herpesvirus)	HHV-8 (KSHV)	141	54	Russo *et al.* (1996); U75698 Neipel *et al.* (1997); U93872
rhesus rhadinovirus (cercopithecoid HV-17)	RRV	131	52	Searles *et al.* (1999); AF083501 Alexander *et al.* (2000); AF210726
mouse HV-68 (murine HV-4)	MHV-4	118	46	Virgin *et al.* (1997); U97553 Nash *et al.* (this issue); AF105037
Herpesvirus ateles (ateline HV-3)	HVA	108	37	Albrecht (2000); AF083424
Herpesvirus saimiri (saimiriine HV-2)	HVS	113	35	Albrecht *et al.* (1992); X64346
equine HV-2	EHV-2	149	58	Telford *et al.* (1995); U20824
wildebeest herpesvirus (alcelaphine HV-1)	AHV-1	131	46	Ensser *et al.* (1997); AF005370

[a] For greatest intelligibility, common and available systematic names are both listed where appropriate.
[b] Genome sizes and (G + C) contents (rounded to whole numbers in kilobase pairs and percentages, respectively) are for the unique sections of the genomes and exclude terminal repeats.
[c] The EBV genome data include only one whole copy of the major repeat array plus the Raji strain version of the 11.8 kbp region deleted in strain B95-8 (Parker *et al.* 1990).

Table 2. *Additional sequenced γ-herpesvirus genes used for phylogenetic trees*

virus species[a]	virus name abbreviation	sequenced genes[b]	references and accession numbers
bovine herpesvirus 4	BHV-4	*08*, *29*	Goltz *et al.* (1994); Z15044 Broll *et al.* (1999); AF139096
cotton-tail rabbit herpesvirus (leporine HV-1)	CRHV	*46*	L33971
equine herpesvirus 5	EHV-5	*08*	Holloway *et al.* (1999); AF050671
Herpesvirus papio (cercopithecoid HV-12)	HVP	*46*	Yates *et al.* (1996); U23857
porcine lymphotropic HV-1	PLH-1	*09*	Ulrich *et al.* (1999); AF191042
porcine lymphotropic HV-2	PLH-2	*09*	Ulrich *et al.* (1999); AF191043
retroperitoneal fibromatosis herpesvirus of macaques	RFHV	*09*	Schultz *et al.* (2000); AF204166

[a] For greatest intelligibility, common and available systematic names are both listed where appropriate.
[b] See § 2 for gene nomenclature.

genomic layout and the clustering of their amino-acid sequence characteristics relative to their counterparts in the other two subfamilies, and the γ-specific genes by their genomic locations and occurrence in all or some of the subfamily. In functional terms, genes from the γ-specific set are seen as being of primary importance in imparting the special characteristics of the γ-*Herpesvirinae*, notably lymphotropisms and neoplastic associations. For the purposes of this paper, the core genes are appropriate for evaluating phylogeny, and otherwise form the background against which the complexities of evolution of the γ-specific genes are examined.

2. PHYLOGENETIC RELATIONSHIPS AMONG THE γ-*HERPESVIRINAE*

We have recently completed an in-depth investigation of phylogeny for the whole family of mammalian and avian herpesviruses based on analysing the relationships among amino-acid sequences for orthologous sets of core genes by maximum-likelihood methods (McGeoch *et al.* 2000). Eight out of the core set of genes are sufficiently well conserved over the whole family to allow adequate amino-acid sequence alignments for use in phylogenetic analysis (McGeoch & Cook, 1994; McGeoch *et al.* 1995); in the usual γ_2 nomenclature these are genes *06* (encoding the major DNA binding protein), *07* (DNA-packaging function), *08* (virion glycoprotein B), *09* (catalytic subunit of DNA polymerase), *25* (major capsid protein), *29* (DNA-packaging protein), *44* (DNA helicase) and *46* (uracil-DNA glycosylase). Complete sequence sets for these genes were available for 19 herpesviruses, including the eight γ-herpesviruses. The underlying assumption was that these eight genes comprise a sample of the conserved elements in herpesvirus genomes sufficiently large and representative that analysis of differences in amino-acid sequences should allow construction of a valid picture of phylogenetic relationships. Here I present the results for the γ-*Herpesvirinae* more fully than previously (McGeoch *et al.* 2000). Sequence alignments were made by an

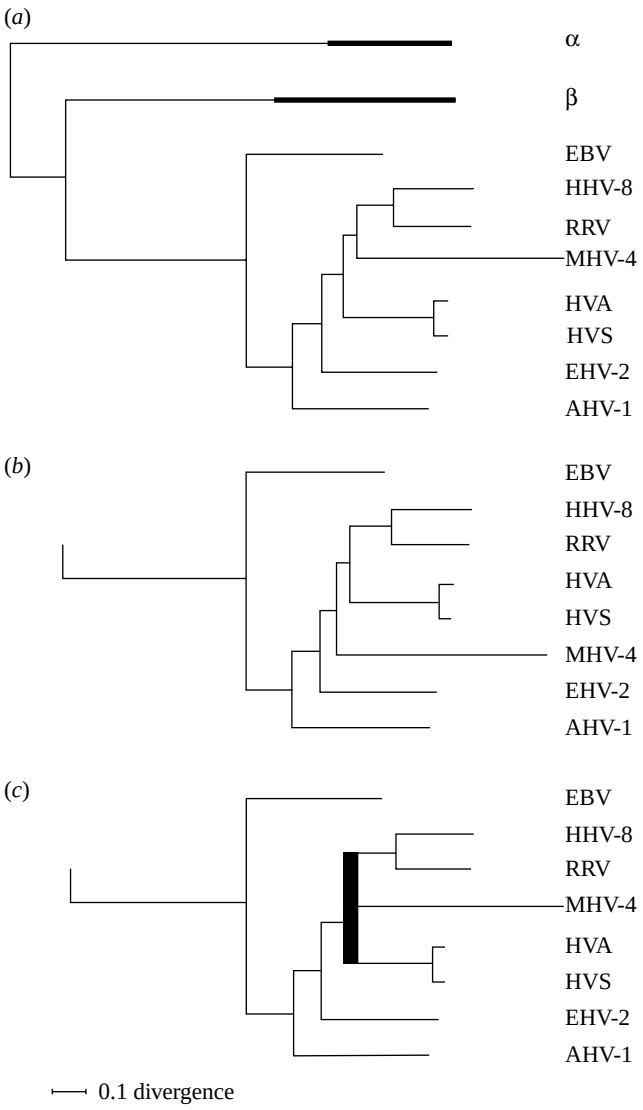

Figure 1. Phylogenetic trees for the γ-*Herpesvirinae* based on eight-gene alignments. (*a*) The top-scoring tree obtained by a maximum-likelihood method (Codeml) with rates for amino-acid sequences of each virus assigned to eight rate classes. The tree's root was taken as the midpoint between the mean locus of branch tips for the α-herpesviruses and the mean locus of branch tips for the β- plus γ-herpesviruses. The α- and β-herpesvirus branches (for 11 species in all) are each represented as compressed to a single heavy line. (*b*) The γ-herpesvirus portion of the second top tree from the same process. (*c*) The γ-herpesvirus section from (*a*) with uncertain branching order around MHV-4 accommodated by drawing as a multifurcation (heavy bar). The divergence scale for all panels is at the foot. (Redrawn from data of McGeoch *et al*. 2000.)

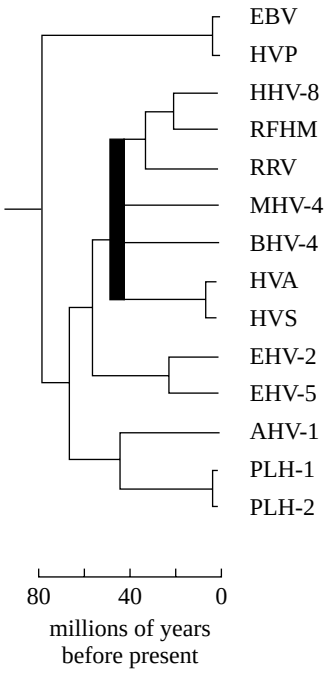

Figure 2. Composite phylogenetic tree for the γ-*Herpesvirinae*. Six additional virus species were added to the tree shown in figure 1*c* from trees based on smaller sets of genes, as listed in table 2. Top-scoring trees were converted to a form that imposed a constant molecular clock to facilitate interpolation of the additional species. The time-scale is based on a value of 47.6 Myr before present for the divergence of Old World and New World primate viruses (Kumar & Hedges 1998). (Redrawn from results of McGeoch *et al*. 2000.)

elaborate procedure using three separate alignment programs and eliminating regions where the programs returned discordant results. When HVA and HVS were lumped together as one operational taxonomic unit then the total number of possible different phylogenetic trees for the resulting seven γ-herpesvirus taxonomic units was 10 395. It was feasible to evaluate these trees exhaustively (while holding topologies for α- and β-herpesviruses constant) using the program Protml (Adachi & Hasegawa 1994) and a set of amino-acid substitution probabilities (Jones *et al*. 1992) to calculate a log-likelihood value for each tree. This exercise was carried out for each of the sets of eight genes and a high level of concordance in top-scoring topologies was found, excepting the locus for MHV-4 which was notably variable. It was then judged reasonable to examine the genes together as a concatenated alignment of 4792 residues, and this was again carried through exhaustively with Protml and then for the top 17 Protml trees using the program Codeml (Yang 1997) for a more elaborate and computationally intensive analysis, with relative rates of substitution at each alignment site for a given virus assigned to a range of eight distinct values by the program. This gave the most broadly based evaluation of γ-herpesvirus relationships accessible, by a refined modern computational modelling procedure.

Results of the analysis are illustrated in figure 1. The top tree (figure 1*a*) scored just marginally better than the next tree (figure 1*b*), which differs only in the locus of MHV-4. Thus even with the large data set employed, a single, unambiguously best tree was not obtained, and the most appropriate achievable representation is that of figure 1*c*, where the branching order of MHV-4 and neighbouring groups is shown as unresolved. When the analysis was repeated omitting MHV-4, a single, clear, top-scoring tree was obtained with equivalent topology to that of figure 1*c*. It may be that resolution could be improved by restricting the virus species in the analysis to

the eight completely sequenced γ-herpesviruses only and adding data for as many more of their genes as possible, but this has not yet been attempted. Smaller sets of whole gene sequences were available for seven additional virus species (table 2) and trees including these data were also evaluated. Loci for six of these further species are shown in the composite tree of figure 2, which was constructed under the constraint of a constant molecular clock in order to facilitate interpolation of the new species' branches (so the tips of branches are in register in this tree). One of the viruses examined, CRHV, was not placed unambiguously with the data available and is not included in the tree; it appeared to belong to the clade comprising HHV-8 through to HVS in figure 2.

The γ-herpesvirus trees of figures 1 and 2 have some complexities for interpretation, more so than seen in the α- and β-herpesvirus subfamilies (McGeoch *et al.* 2000). MHV-4 shows an uniquely long terminal branch—apparently this lineage has been evolving at an enhanced rate relative to others, and it is attractive to associate this behaviour with the inability to place MHV-4 precisely on the tree. Conversely, the EBV lineage may have evolved unusually slowly. Next, within the trees for the α- and β-herpesviruses there are several sections where congruence of the branching patterns with those of the corresponding host lineages powerfully indicates a co-speciational origin for those virus lineages and the relative magnitudes of substitutional divergences are also consistent with co-speciation (McGeoch *et al.* 2000). In the γ-herpesvirus tree this trend is discernible but less complete. Thus, branching of the Old World lineage of primate viruses (HHV-8, RRV, RFHV) from that of the New World (HVA, HVS) could be co-speciational, but the extent of their divergence would then indicate an overall twofold elevation in substitution rate compared with the α- and β-herpesvirus examples. Also, the ungulate virus lineages of AHV-1 and EHV-2 do not form a clade (which would indicate co-speciation), unlike the ungulate virus lineages in the α-herpesviruses. I register a residual uneasiness with this aspect of the present γ-herpesvirus tree—despite the size of the data set and the quality of the procedures used to derive figure 1*c*, I believe it marginally possible that factors such as differential evolutionary rates might have acted to confound inference of ungulate virus branching pattern.

Notwithstanding these complications, the observed co-speciational correspondences and the extent of sequence divergences as well as comparison with the cognate α- and β-herpesvirus trees, all point to the γ-herpesvirus tree representing a great depth of evolutionary time. I have therefore applied a time-scale to figure 2 based on taking the branch point of the Old and New World primate virus lineages as equivalent minimally to the divergence time of the host lineages, recently estimated as 47.6 Myr before present from DNA sequence data (Kumar & Hedges 1998) and representing a revision of the value of 35 Myr based solely on palaeontology and used previously (McGeoch & Cook 1994). This corresponds to an overall evolutionary rate of 6×10^{-9} substitutions per site per year in each lineage. This coarse estimate should be helpful for considering long-range evolution, but I emphasize that both stochastic and systematic errors must be considered to apply.

There are several instances in the γ-herpesvirus tree of more than one closely related virus species associated with a given host (in porcine, equine and macaque viruses), a phenomenon seen also in the α- and β-herpesviruses. Current research is expanding the catalogue of Old World primate γ_2 viruses, primarily by PCR methods. Sequences of two closely related viruses in the RFHV lineage with distinct macaque host species have been described, and also another macaque virus related to RRV (Schultz *et al.* 2000), and sequences have been recovered from African green monkey tissues that evidently represent two novel γ_2 viruses, one related to the RFHV lineage and the other to RRV (Greensill *et al.* 2000). HHV-8-related sequences have also been detected in gorillas and chimpanzees (Lacoste *et al.* 2000). It thus appears likely that these two lineages—RRV-like and RFHV-like—are each populated by viruses from a range of Old World primate species (Damania & Desrosiers, this issue), with HHV-8 being the human member of the RFHV group, so there might exist a presently undetected human herpesvirus that belongs to the RRV group—a possibility of potential significance for human medicine. Such close, parallel lineages could arise in a variety of ways—for instance, through geographical partition of a host species, via different tissue tropisms of virus species, and by transfer between host species. In general, the phenomenon of closely related virus lineages in closely related host species urges that it would be prudent to reserve arguments about possible co-speciation for cases involving higher host taxa and larger viral divergences—among the γ-herpesviruses, the Old World–New World primate virus split might be taken as a minimal acceptable instance.

Whatever the underlying detail of branching for the MHV-4 lineage (and the associated BHV-4; see figure 2), these clearly have not arisen by a co-speciational route; interestingly, bovine herpesvirus 2 occupies an equivalent locus in the α_1 group between lines of Old World and New World primate viruses (McGeoch & Cook 1994). Reasons for the refractory behaviour of MHV-4 sequence data in phylogeny programs have not really been resolved and would bear further investigation—variability in inferred locus appears a rather general property for distinct subsets of data (therefore discounting recombination between two parental virus species as the underlying cause), and the apparently higher evolutionary rate seems insufficient to give rise to such problems (and indeed it might be another manifestation rather than the cause).

Limited PCR-derived sequence data are also expanding membership of the AHV-1 clade, with addition of ovine herpesvirus 2 and of a previously unknown bovine lymphotropic virus (Rovnak *et al.* 1998); this clade is therefore emerging as populated with viruses from a significant range of artiodactyl species. The same approach has added additional perissodactyl viruses to the EHV-2–EHV-5 clade (Ehlers *et al.* 1999), and these reinforcements again focus attention on the nature of the relationship between these two ungulate virus clades. In this connection, another complexity is that EHV-2 and EHV-5 are unique among characterized γ-herpesviruses in that they do not possess at their genome termini sets of tandem reiterations of approximately 1 kbp: EHV-2 has single 17.5 kbp direct repeats at its genome extremities

Table 3. *γ-Herpesvirus-specific genes and cellular homologues*

(References are given in tables 1 and 2 and in the text.)

cellular gene or function	virus species	virus gene(s)
G-protein-coupled receptor[a]	EHV-1	*E1*
G-protein-coupled receptor[a]	AHV-1, EHV-1	*A5*, *E6*
G-protein-coupled receptor (interleukin 8 receptor)[a]	EHV-2, HHV-8, HVA, HVS, MHV-4, RRV	*74*
dihydrofolate reductase (DHFR)	HHV-8, HVS, RRV	*02*
α-N-formylglycineamide ribonucleotide aminotransferase (FGARAT)	all	*03*, *75*, *BNRF1*
serpin	MHV-4	*M1*
semaphorin	AHV-1	*A3*
complement control protein	HHV-8, HVA, HVS, MHV-4, RRV	*04*
interleukin 6 (IL-6)	HHV-8, RRV	*K2*, *R2*
CC chemokines	HHV-8, RRV	*K4*, *K4.1*, *K6*, *R4*
CD59	HVS	*15*
bcl-2	AHV-1, EBV, HHV-8, HVS, HVA, MHV-4	*A9*, *BHRF1*, *16*, *M11*
interferon regulatory factors (IRFs)	HHV-8, RRV	*K9*, *K10*, *K10.1*, *K11*, *R9.1–R9.8*
thymidylate synthase (TS)	EHV-2, HHV-8, HVS, HVA, RRV	*70*
interleukin 10 (IL-10)	EHV-2, EBV, HVP	*E7*, *BCRF1*
apoptotic protease inhibitor	EHV-2, HHV-8, HVS, HVA, RRV	*E8*, *K13*, *71*, *R13*
cyclin D	HHV-8, HVS, HVA, MHV-4, RRV	*72*
OX-2 cell surface protein	HHV-8, RRV	*K14*, *R14*
caspase recruiting protein	EHV-2	*E10*
β-1,6-N-acetylglucosaminyltransferase	BHV-4	*BORFF3–4*

[a] The three classes of G-protein-coupled receptors listed are presumed distinct, and EHV-2 *E6* and AHV-1 *A5* may not be orthologues.

(Telford *et al.* 1995) and EHV-5 has no substantial repeat structures (Agius *et al.* 1992).

The substantial enlargement of the γ_2 group of herpesviruses via DNA sequence data, in many cases for otherwise uncharacterized viruses, is now making the current taxonomic description of the *γ-Herpesvirinae* (see Minson *et al.* 2000) look in need of revision. My personal view is that a more extensive framework is needed to accommodate the range of viruses detected, and that the phylogenetic trees described in this section could provide a robust foundation for such an enlarged taxonomic scheme. By comparison, the γ_1 lineage now appears sparsely populated indeed, with only a handful of Old World primate viruses, although this is clearly about to change with the very recent discovery of a New World γ_1 virus from lymphoproliferative disease in a common marmoset (Ramer *et al.* 2000; Wang *et al.*, this issue). The trees of figure 1 do somewhat 'undersell' the γ_1 lineage, and it should be emphasized that it is not merely another clade attached to the populous γ_2 tree, but comprises a very distinct group in respect of its unique complement of latent-cycle genes.

3. OCCURRENCE AND RELATIONSHIPS OF γ-SPECIFIC GENES

γ-Herpesviruses have a variable complement of γ-specific genes, some present in multiple virus species, others in one only. These include genes well studied in their own right, genes with non-herpesvirus homologues and genes that are still hypothetical or have no indication of function. By current criteria there are eight genes that are γ-specific and are found in all of the sequenced genomes. However, these are of low value for interpretive purposes inasmuch as little is known about their functions, aside from the transcriptional activator encoded by EBV gene *BRLF1*, and its γ_2 gene *50* equivalents. Approximately 18 distinct classes of cellular genes or gene families of known function have been found that have γ-herpesvirus homologues, all but one of which are found only in subsets of the viruses. These genes, listed in table 3, include a number that might readily act to alter cell state or behaviour, and which could thus play a role in evolution of specific γ-herpesvirus strategies. Several have been shown to have cell transforming potential—in terms of numbers of candidate oncogenes, HHV-8 is particularly striking (reviewed by Neipel & Fleckenstein 1999; Moore & Chang, this issue).

The γ-specific genes are very strikingly concentrated in just a few genomic loci. Just how many such regions are defined depends on how tightly they are demarcated. In a discerning study of γ-herpesvirus variation, Nicholas *et al.* (1998) delineated eight major divergent loci. For the purposes of this paper I regard the regions adjacent to the two genomic termini as of greatest importance, representing loci A plus B of Nicholas *et al.* (1998) at the left end of unique sequence in the genomes (in the conventional γ_2 orientation) and E plus F at the right end. Figure 3 depicts the layout of sequence elements in the unique region adjacent to the left terminus of each of the sequenced γ-herpesvirus genomes, from the left extremity of unique sequence to the start of the block of core genes *17* through to *50*, and in the unique sequence adjacent to the right genomic terminus, from the right of the block of

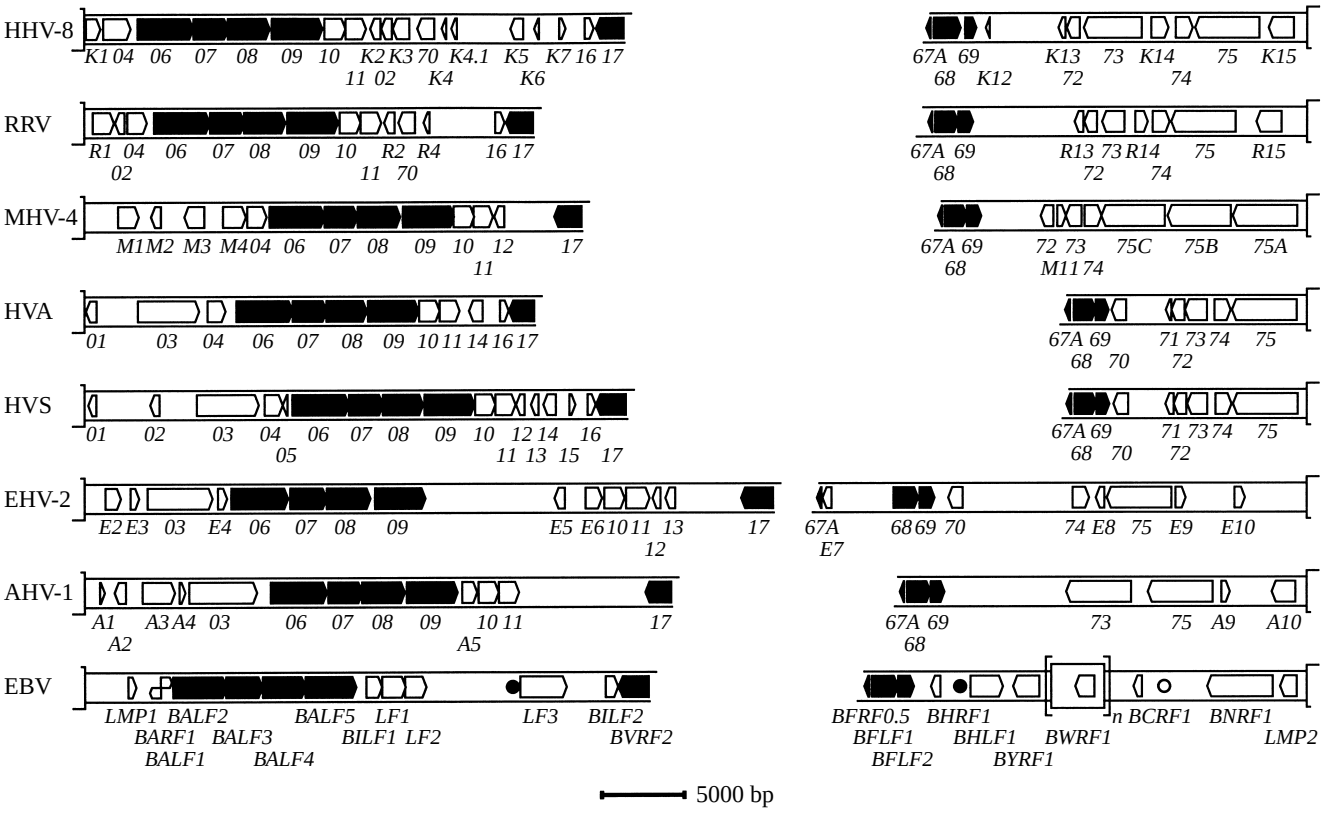

Figure 3. Gene layout of γ-herpesviruses in regions adjacent to the genomic termini. Unique sequence regions adjacent to the left and right termini of the eight sequenced γ-herpesvirus genomes are depicted, from the left extremity to gene *17* (*BVRF2* in EBV), and from gene *67A* (*BFRF0.5* in EBV) to the right extremity. Coding regions are depicted for core genes as black arrowheads and for other genes as open arrowheads. Details of exons in genes with multiple-exon coding regions are not shown. The EBV genome is inverted in this view from its conventional orientation, and only one copy of the major internal repeat sequence is shown (large box); this is actually present as multiple copies ($n = 12$ in the genome sequence library entry). In the EBV genome the two copies of *oriLyt* are marked by filled circles and *oriP* by an open circle. Length scale is indicated at the foot.

core genes *58* to *67A* through to the right extremity of unique sequence. The sections shown range from 27 to 42 kbp for the left section and from 14 to 30 kbp for the right, and it is apparent that gene content is highly diverse. There are also other classes of elements present: some viruses, including EBV, HVS, HVA and MHV-4, express non-protein-coding RNAs in these regions, and there are substantial sections in certain genomes (e.g. EHV-2, AHV-1) that have no assigned function. In EBV there are origins of DNA replication (*oriLyt*) located in both terminal sections and also a latent-cycle origin (*oriP*) in the right section (see figure 3, closed and open circles, respectively). From analysis of strand-specific bias in nucleotide composition, occurrence of tandem reiterations and other sequence characteristics it seems likely that the γ_2 viruses also have origins in approximately equivalent locations to both *oriLyt* copies of EBV.

My intent in this section is to employ features of the occurrence and relationships of appropriate γ-specific genes to make inferences about their evolutionary histories. The primary approach used is evaluation of amino-acid sequence alignments for the virus genes and available homologues from potential host lineages through construction of gene trees. In principle, for a given set of aligned sequences one might obtain information on the likely origin of a virus gene as a lineage branching from the host tree, on extent of its divergence since entering the virus lineage, and on relationships among the virus members. In practice, the exercise can be restricted by availability of appropriate host gene sequences or by the sequence set under study being insufficiently informative to give adequate resolution—there are also limitations and pitfalls in the tree modelling algorithms available. The leftmost and rightmost genes (HHV-8 K1 and K15) in the Old World primate γ_2 viruses are separately of interest for their contemporary variation and are discussed in § 4. Some degree of insight is provided by eight sets of genes, namely (i) IL-10 genes in γ_1 viruses (gene *BCRF1*) and in EHV-2 (E7); (ii) the gene *02* set encoding DHFR; (iii) genes *03* and *75*; (iv) the IL-6 (*K2*) genes; (v) the gene *70* set encoding TS; (vi) genes *12*, *K3* and *K5*; (vii) genes *K4*, *K4.1* and *K6* encoding CC chemokines; and (viii) the gene *74* set, as will be described at varying length (abbreviations are given in table 3). The other gene sets in figure 3 are not dealt with individually. For all of the gene sets in figure 3 a plausible pattern of acquisition and loss can be discerned that is compatible with the branching pattern of the phylogenetic tree described in § 2, although these data are not themselves sufficient to predict phylogeny.

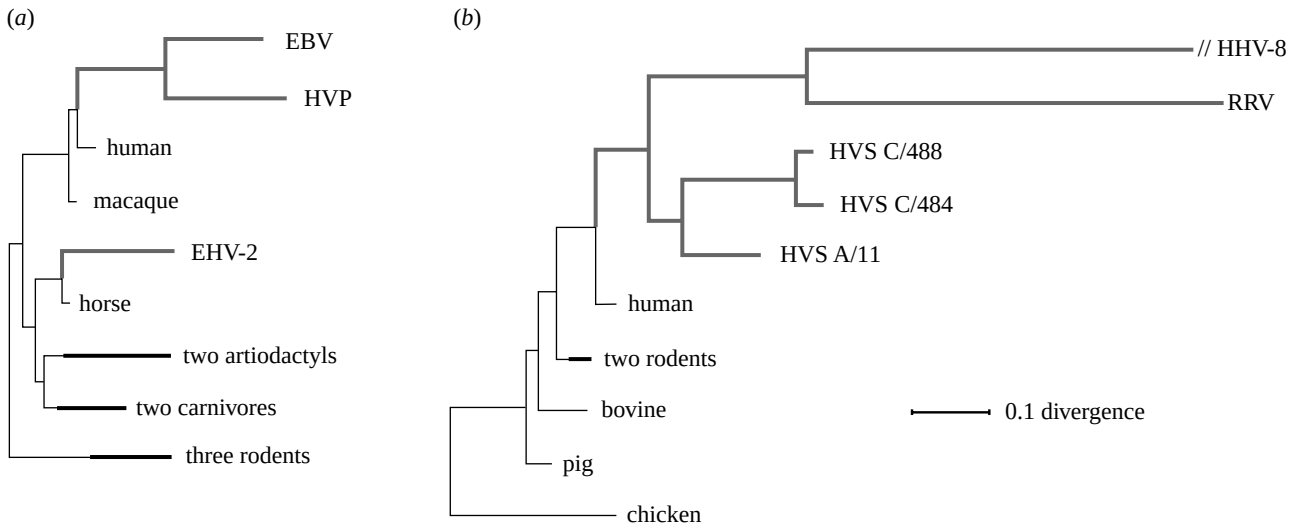

Figure 4. Gene trees for virus and cellular versions of IL-10 and DHFR. (*a*) Amino-acid sequences for ten mammalian IL-10 proteins were chosen as a useful tree-building set from a larger number in the Owl database and, together with EBV, HVP and EHV-2 IL-10 sequences, gave a high quality alignment of 166 residues excluding positions with a gap in any sequence. A tree was derived using Protml and Codeml, as described in § 2 for phylogenetic tree construction. The tree as shown was rooted at its midpoint, after excluding the virus sequences. Virus-specific branches are shown in grey, and branches for artiodactyl, carnivore and rodent sequences are drawn in compressed form. (*b*) Similarly, six cellular DHFR sequences and γ-herpesvirus DHFR sequences for HHV-8, RRV and three HVS strains gave an alignment length of 184 residues and the tree shown was derived and rooted in the same manner. The HHV-8 branch is drawn truncated by half. The scale bar at lower right applies to both trees.

(a) *Viral IL-10 genes*

EBV and its baboon counterpart HVP (Yates *et al.* 1996) have equivalently located homologues of the IL-10 gene (*BCRF1*) while EHV-2 alone among the γ_2 viruses has an IL-10 gene (*E7*), at a distinct genomic location (figure 3). IL-10 sequences are known for many mammalian species, and a good quality gene tree was derived based on the aligned amino-acid sequences (figure 4*a*). The IL-10 genes of the two γ_1 viruses are seen to have a common origin from a locus in the primate lineage close to the divergence point of the human and macaque genes, and the EHV-2 gene has a distinct origin with the horse sequence as its nearest cellular counterpart. Clearly the γ_1 IL-10 genes represent a single capture event from a primate genome, and the EHV-2 gene a separate capture from an equine genome. In both instances the viral versions have diverged significantly, although overall sequence similarity remains high, and the relative branch lengths for the viruses and the host genes indicate that the amino-acid sequences of the viral IL-10s have been changing an order of magnitude faster than those of the host. Based on estimates of non-synonymous divergence of virus and host sequences, the capture events both occurred around 10^7 years ago.

(b) *Viral DHFR genes*

The distribution and relationships of the gene *02* set, encoding DHFR, in primate γ_2 genomes present a case of some intricacy. Versions of the gene are present in HHV-8, RRV and HVS, but not in HVA or non-primate viruses. The HVS and RRV genes are at equivalent genomic loci and the HHV-8 gene is at a distinct locus (figure 3). Homologous genes with well-conserved amino-acid sequences are known for both eukaryotes and prokaryotes. Figure 4*b* shows a gene tree for available eukaryotic and herpesviral DHFRs. The viral DHFRs form a single clade associated with the primate host branch (as represented by the human sequence), consistent with the gene having entered the primate γ_2 virus genome from the primate lineage at an early point in this virus lineage. The HHV-8 and RRV genes form a sister clade to that of the three HVS strains available; thus, the viral DHFR gene tree matches the branching pattern in the broader phylogenetic tree of figure 1*c*. The most economical inference is that the HVS–RRV location is ancestral and the location of the DHFR gene in the HHV-8 genome has changed since divergence of the HHV-8 and RRV lines. Most striking is the large extension of the HHV-8–RRV specific branches relative to their HVS equivalents, a fivefold increase in mean length—apparently the DHFR genes in the Old World primate viruses have experienced a process of enhanced substitution. Thus, although the RRV and HVS genes occupy equivalent loci and the RRV DHFR amino-acid sequences are more similar to those of HVS than of HHV-8, in a phylogenetic sense HHV-8 DHFR is the closest relative of RRV DHFR since they had the most recent common ancestor.

These interpretations are crucially dependent on the correctness of the inferred tree topology. It should be noted that connections made by very long branches represent a notoriously artefact-prone aspect of tree building—indeed in a previous pass at this topic using the same data, except that the RRV sequence was not then known, we obtained a tree with the HHV-8 branch located at a distinct locus deeper into the mammalian gene tree (McGeoch & Davison 1999*a*). Thus, what assurance is there of the present account's reliability? First, the result for viral DHFR genes emerged strongly supported from an extensive process of tree evaluation. Second, the tree is

otherwise well behaved in that the host portions exhibit an approximation of expected relationships (although resolution of deep branches is limited with this highly conserved data set). Third, the branching pattern of the virus genes gives a believable origin for them from the primate lineage, and constitutes a maximally parsimonious explanation of the genomic distribution of viral DHFR genes, namely a single capture event from the host then subsequent relocation in HHV-8 and loss in HVA. Finally, the HHV-8 and RRV DHFR genes are seen to behave consistently in their high substitution rates, and this phenomenon has precedents in the rate variations inferred for the eight-gene phylogenetic tree.

(c) *Genes* 03 *and* 75

Gene *03* is present in all the γ_2 viruses except HHV-8, RRV and MHV-4, and has thus been lost from these latter lines. There is a paralogous gene (*75*) at the other extremity of the genome, which is both ubiquitous within the γ-herpesviruses and specific to them—the *03* and *75* versions are substantially diverged and are presumed to be related by an ancient duplication/translocation event. There is an EBV counterpart, *BNRF1*, at the equivalent genomic location to gene *75*. The EBV and HVS gene products have been reported as components of the virion tegument (Cameron *et al*. 1987). The amino-acid sequences are pronouncedly similar to that of α-N-formylglycineamide ribonucleotide aminotransferase (FGARAT), a cellular enzyme of the purine biosynthetic pathway—I am not aware of any study of possible enzyme activity of a viral version, and the relationship presents as somewhat of a mystery. The *03* and *75* genes belong to a class, including also gene *70* and gene *16*/*M11*/*BHRF1*, whose location varies between the two terminal regions. Interestingly, in MHV-4 there are three contiguous, diverged versions (genes *75a*, *75b* and *75c*)—perhaps compensating for loss in the MHV-4 line of gene *03*.

(d) *Viral IL-6 genes*

Gene *K2* in HHV-8 and its counterpart in RRV encode a homologue of IL-6. We previously showed (McGeoch & Davison 1999*a*) that the lineages of the two viral genes appear to originate as a single clade from deep in the lineage of primate IL-6 genes, as with the DHFR gene discussed above, and like the HHV-8 and RRV DHFR genes, the viral IL-6 genes have both diverged greatly from their host gene counterparts.

(e) *Viral TS genes*

A subset of γ-herpesviruses and one other member of the herpesvirus family, the α-herpesvirus varicella-zoster virus (Davison & Scott 1986), possess TS genes. HHV-8 and RRV have TS genes in a location near the left end of the genome, while in HVA, HVS and EHV-2 the TS gene (*70*) is located near the right end of the genome. We previously analysed relationships among these viral TS genes and cellular homologues (McGeoch & Davison 1999*a*). The TS protein sequences, viral and eukaryote, are so highly conserved that their alignment does not yield sufficient information for resolving deep parts of the gene tree. The HHV-8–RRV pair and the HVA–HVS pair are each relatively close, but more distant relationships were not analysable. The distribution among virus species suggests that the right end locus is ancestral, so the HHV-8–RRV version might represent a translocation event between the two genomic terminal neighbourhoods (as mentioned above for genes *03*/*75* and gene *16*) or alternatively an independent capture from the host genome.

(f) *Genes* 12, K3 *and* K5

HHV-8 genes *K3* and *K5* are homologous and were originally regarded as related to a BHV4 gene but lacking other γ-herpesvirus homologues (Russo *et al*. 1996; Van Santen 1991). In fact they are also weakly but definitely related to HVS gene *12* and to an MHV-4 gene. EHV-2 gene *12* is probably unrelated, while the absence of RRV and HVA homologues must represent separate losses from their lines. The *K3* and *K5* genes have recently been shown to be involved in viral evasion of cellular immune responses (Ishido *et al*. 2000*a*,*b*).

(g) *CC chemokine genes*

HHV-8 has three CC chemokine genes, *K4*, *K4.1* and *K6*. We previously concluded that *K4* and *K6* probably had a common origin, from a presently unidentified host homologue, and that *K4.1* may have represented a separate acquisition from the host genome (McGeoch & Davison 1999*a*). The analysis was limited by the short length of the amino-acid sequence alignment (77 usable residues) and complicated by the great numbers of closely related mammalian genes in this class. Subsequently the RRV sequence became available. RRV has only one CC chemokine gene, which is probably primarily related to the HHV-8 *K4*–*K6* pair, although it is significantly diverged—this result also served to emphasize that derivation of HHV-8 *K4.1* from *K4* could be a credible alternative scenario. Gene *K5* (discussed above) and gene *K6* are adjacent, and genes *K3* (paralogous to *K5*) and *K4* are separated only by the TS gene (figure 3), so that there are two portions of the HHV-8 genome (*K3*–*TS*–*K4* and *K5*–*K6*) that are almost adjacent and that each contain two separate homologous elements. The coding sequences of *K4* and *K6* are sufficiently close that a value for synonymous substitution divergence could be obtained (*ca*. 1.5 substitutions per site), which suggests a likely divergence time in the virus genome of the order of 10^7 years ago for these two genes, and a similar result holds for *K3* and *K5*. The HHV-8 lineage has thus experienced a local duplication, probably postdating divergence of the HHV-8 and RRV lines, and subsequently partly overlaid by other genomic changes.

(h) *Gene* 74

Gene *74*, found in six γ_2 viruses, encodes a protein with seven transmembrane segments, with homology to G-protein-coupled receptors and to IL-8 receptors in particular. Although the virus genes are strongly diverged, in tree building exercises they form a clade with branching order equivalent to that of figure 1*c*. Their relationship to the host homologues remains ill-defined: the host IL-8 A- and B-class receptors are highly conserved, forming a very compact tree, and all that can be said is that the virus clade joins the host tree at an unresolved point in its interior.

(i) *Other γ-specific genes*

Members of the gene *16*, gene *71*, gene *72* and gene *K14* sets present similar cases to that described for gene *74* sequences: in each case the virus sequences are strongly diverged and have detectable host counterparts, typically involving host multigene families, and in each case the relationship with the host sequences is too distant to ascribe a specific host lineage as the origin of the virus genes. In some cases (e.g. HVS gene *71* and EHV-2 gene *E8*), virus genes are so diverged that they are almost 'over the horizon' from each other, although both are readily recognized as related to the same host homologue. In addition to the cases discussed above, there are two notable exceptions to this generality of great divergence. First, EHV-2 E10 uniquely contains an N-proximal caspase recruiting domain with a closely similar cellular homologue, although the two proteins are distinct in their C-terminal regions (Koseki *et al.* 1999; Thome *et al.* 1999). Second, BHV-4 uniquely encodes a β-1,6-N-acetylglucosaminyltransferase with close cellular counterparts (Vanderplasschen *et al.* 2000). Finally, outside the genomic terminal regions, the sets of IRF-related genes seen only in HHV-8 and RRV present an interesting case. In HHV-8 there are four of these in tandem (*K9*, *K10*, *K10.1* and *K11*), and in RRV there are eight genes in tandem (*R9.1* to *R9.8*) (Searles *et al.* 1999; Alexander *et al.* 2000). *R1* to *R4* are most similar to *R5* to *R8*, respectively, so evidently the RRV set has most recently expanded by a straightforward duplication from four to eight genes. The genes in both viruses are all well diverged from the host version and also from each other, so might have evolved distinct capabilities.

(j) *EBV genes*

Although this section was written to include EBV as well as the γ_2 herpesviruses, EBV really has no visible commonality of host gene homologues with the other viruses beyond that of *03/75/BNRF1* with FGARAT, *16/M11/BHRF1* with the bcl-2 family, and the independently acquired IL-10 homologues of EBV and EHV-2. EBV and its primate virus relatives have a well-studied set of genes concerned with host cell transformation and maintenance of the virus genome in dividing cells. At the left terminal region (drawn as in figure 3, which inverts the EBV genome from its usual presentation), there is the γ_1-specific *LMP-1* gene. The right terminal region contains the *LMP-2*, *LP* and *EBNA-2* genes; the latter two are shown in figure 3 as *BWRF1* and *BYRF1*, respectively. *LMP-2* is a definite relative of the *LAMP-2* (*K15*) genes of HHV-8 and RRV, although a distant one. Both *LMP-2* and *LAMP-2* encode proteins with 12 transmembrane segments in genes with multiple exons, although the amino-acid sequences are dissimilar. However, the distribution of the hydrophobic membrane sequences, and separately the lengths of the coding exons 2 through 6 (numbered for the *LMP-2A* transcript version of the EBV gene) correspond closely between EBV and HHV-8 (Hayward 1999). *LMP-2A* and *LAMP-2* differ in that *LMP-2A* possesses an N-terminal cytoplasmic domain with sites for signal transduction activities, while *LAMP-2* has instead a generally similar C-terminal domain. The *EBNA-3* family of three distantly related genes is located tandemly at an internal locus in the EBV genome, and the *EBNA-1* gene at another. In latently infected and transformed cells the *LP*, *EBNA-2*, *EBNA-3A*, *EBNA-3B*, *EBNA-3C* and *EBNA-1* genes are expressed by way of several versions of a giant transcript across most of the genome and differential, complex splicing (reviewed by Henderson *et al.* 1994)—this system has no known equivalent in the γ_2 viruses.

The comparative exercises described in this section demonstrated for the viral genes encoding versions of IL-10, DHFR and IL-6 that they are plausibly derived from the host lineages and that the transfer occurred in the deep past. However, with the exception of the IL-10 case, even in these best examples the analysis was stretched by the large divergences seen between virus genes and between virus and host genes. In other cases, relationships with the host genes are detectable but resolution remains low. Factors contributing to these difficulties, in addition to the extensive divergence of the virus sequences and the short lengths of some genes, include in various cases: (i) insufficient numbers of host genes available for comparison; (ii) very high conservation of amino-acid sequence among host genes; and (iii) existence of multiple families of host homologues. The imminent availability of a near complete listing of our human gene set may well impact on resolving which members or subset of multigene families are most likely to be the progenitors of virus homologues, as well as increasing the numbers of virus genes with detected host homologues.

Several general points arise from examination of γ-specific gene complements. The terminal regions of these virus genomes have been the sites of a range of classes of genomic change, including translocation, duplication and deletion as well as capture of host genes. I suggest that such enhanced receptivity to recombinational events in the terminal regions might to some degree be consequent on the presence of elements whose function directly or indirectly involves recombination, namely genomic terminal tandem arrays and origins of DNA replication. I emphasize that the view is of action over a great span of time—from the phylogenetic analysis of core genes, the whole process of divergence of the γ-herpesviruses took around 10^8 years. It is evident from the patterns of occurrence of specific genes (figure 3) that loss of genes has been of the same order of frequency as gain and rearrangement—thus, the evolution of γ-herpesvirus strategies has not been equivalent simply to accumulating an ever larger panoply of genes effective in subverting cellular processes. It is striking that all of the sequenced genomes are distinct in their complements of γ-specific genes; all their strategies for modulating host processes must thus differ to some degree, yet all are survivors of the evolutionary process and as such represent solutions that have enabled long-term coexistence of virus and host.

4. VARIATION IN THE HHV-8 GENOME

In the short span since discovery of HHV-8 in 1994, studies of molecular evolutionary processes in this virus have developed to become the most detailed, interesting and challenging among the human herpesviruses. The basis of this phenomenon lies primarily with attributes of the leftmost and rightmost genes in the HHV-8 genome—at the left extremity, gene *K1*'s sequence has

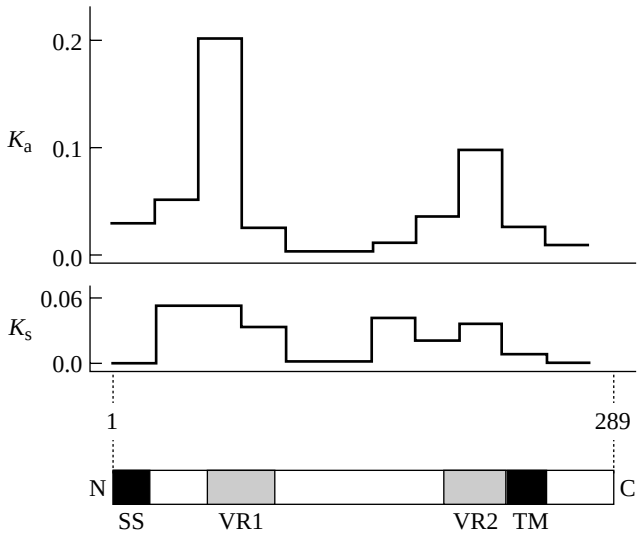

Figure 5. Structure and diversity of the HHV-8 *K1* coding region. The bar at the base of the figure represents the complete amino-acid sequence (residues 1–289) of the *K1* gene product. Black-filled areas SS and TM represent the N-terminal signal sequence and C-proximal transmembrane sequence, respectively, and the two grey areas represent the hypervariable regions VR1 and VR2. In alignment with the amino-acid sequence diagram, and present measures of non-synonymous and synonymous divergence (K_a and K_s, respectively), as follows. For a set of 12 *K1* coding sequences belonging to clade A, all pairwise values for K_a and K_s were computed for successive, non-overlapping regions of 25 codons, and the mean K_a- and K_s-values for each region presented as histograms.

emerged as extraordinarily more diverse than any other gene of HHV-8 or indeed any herpesvirus, while at the right end the *K15* or *LAMP-2* gene occurs as two highly diverged alleles. Analysis of variation in these two genes is of intrinsic interest and in addition their use as molecular epidemiological markers has enabled insights into the molecular evolution of HHV-8 over the last 10^5 years.

The *K1* open reading frame encodes a type I glycoprotein of some 289 residues, which has been demonstrated to possess cell transformation activity (Lee *et al.* 1998); the cytoplasmic domain of the K1 protein has potential binding sites for cell-signalling tyrosine kinases. Sequence analysis of the gene from many sources, mostly via PCR, has revealed differences between pairs of amino-acid sequences ranging up to over 40%, with two particularly variable loci termed VR1 and VR2 (Cook *et al.* 1999; Meng *et al.* 1999; Zong *et al.* 1999). Figure 5 illustrates for a set of *K1* sequences how divergences vary across the gene. Eleven cysteine residues in the extracellular domain remain highly conserved, indicating preservation of the overall fold. Comparison of the two available genomic sequences of HHV-8, which contain *K1* sequences of types A and C (see below), shows that the levels of synonymous and non-synonymous substitution in the *K1* gene are, respectively, one and two orders of magnitude greater than in the rest of the genome (excluding the *K15* gene). In pairwise comparisons of *K1* coding sequences, non-synonymous divergence is always greater than synonymous divergence, on average by a factor of 2. Diversity in *K1* therefore constitutes an unambiguous example of positive selection for change in the amino-acid sequence. Hayward (1999) has pointed out that the range of amino-acid substitution per site observed for VR1 is particularly extensive, with the core ten-residue stretch of the region exhibiting five to ten distinct amino acids at each position. Notably, no differences have been detected in *K1* sequences obtained from distinct sites or at different times in a single infected person (Meng *et al.* 1999; Zong *et al.* 1999)—de-emphasizing comparison between the cases of HHV-8 *K1* and HIV *env*.

In principle, the mechanistic basis of the selection pressure could represent either evasion of a host response by the virus or a requirement for interaction with some host component that is highly variable in human populations. Given the general absence of comparable effects in other HHV-8 genes and other herpesviruses, straightforward selection by way of antibodies or T cells seems rather unlikely, although it is hard to avoid the impression that the major histocompatibility complex (MHC) system is likely to be involved in some manner in the pressure for *K1* diversity. Hayward (1999) aired the idea that perhaps the VR1 region of *K1* is involved in presenting a range of possible HLA class I epitopes, but did not reach a complete rationale for this. It is notable that VR1 varies in sequence but not in length, unlike VR2. Knowledge of the K1 protein's function, of its significance for the viral life cycle and of its interactions with host proteins is still very limited, and this paucity inevitably impacts on our appreciation of the significance of its sequence variation.

I estimate that non-synonymous substitution per site for the most variable region within VR1 is about 200-fold higher than synonymous substitution per site observed for the HHV-8 genome excluding *K1* and *K15*. Thus, if the latter value is taken as approximately equivalent to total substitution mutation per site outside *K1* and *K15*, then the selection process might have accelerated the rate of DNA sequence change in VR2 relative to the bulk of the genome by over two orders of magnitude. How could this have been achieved? Given, first, a strong selection pressure for change in protein sequence targeted primarily at VR1 and VR2 and, second, action over an extended period of time, I believe that the main features of the process are explicable by application of established principles and without invoking any *ad hoc* special mechanism (such as local error-prone replication of DNA). I suggest that high relative rate of substitution in *K1* can be understood in terms of the hitchhiking effect, whose basis is that positive selection acting on a sequence variant at a particular genomic locus will also necessarily select the version of overall genomic sequence possessed by the DNA molecule carrying the variant locus (or, more precisely, by that length of DNA that is not recombinationally decoupled from the variant site) and diversity in the progeny population of genomes will thereby be reduced (Kaplan *et al.* 1989). Thus I am proposing that the amount of sequence change observed in VR1 and VR2 represents that proportion of the total amount of mutation that is selected under the postulated pressure, while the amount of change seen elsewhere in the genome has been reduced from the mutational total both by normal purifying selection for each gene and also by linkage to the *K1* selection process. This suggestion

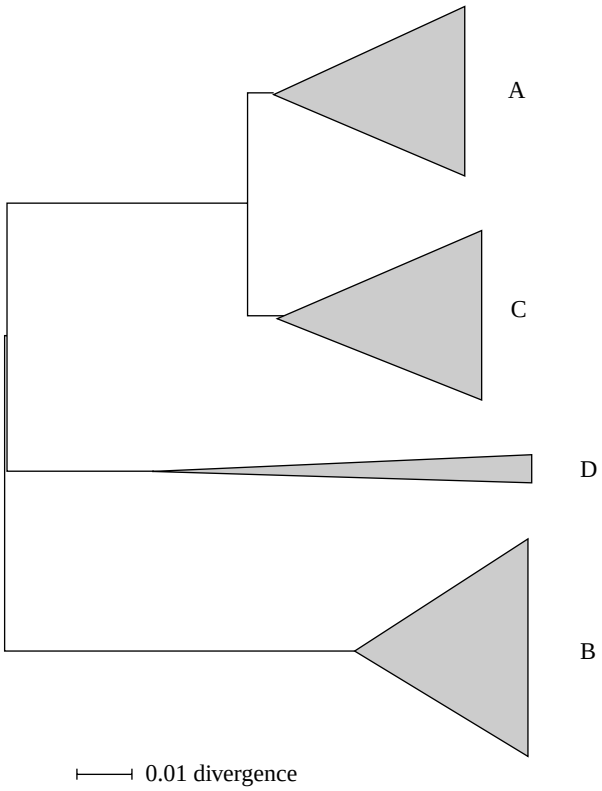

Figure 6. Clades of HHV-8 *K1*. A summary tree is depicted based on a neighbour-joining analysis of 45 complete, aligned *K1*-coding DNA sequences. The tree was rooted at its midpoint. Branching within clades A, B, C and D (Hayward 1999; Cook *et al*. 1999) is not shown in detail but as shaded areas with their bases proportional to the number of sequences in each clade. The divergence scale is shown at the foot.

constitutes only the bones of a possible mechanism. There are complexities in the *K1* data that need to be accounted for, namely the occurrence of high synonymous substitution in VR1 and VR2 and the occurrence of relatively high non-synonymous substitution within *K1* but outside VR1 and VR2. The first of these probably arises in part from the fact that the genetic code provides various pathways by which several successive non-synonymous changes at one codon can result overall in an apparent single synonymous change. I note that it may be feasible to model *K1* variation quantitatively using the coalescent theory of genealogical change, to obtain improved insight into the elements of the process (Kaplan *et al*. 1988; Hudson 1990).

K15, the rightmost gene in the HHV-8 genome, has eight coding exons which specify a protein with 12 membrane-spanning regions and a C-terminal cytoplasmic tail thought to carry binding sites for signalling proteins. The function of the K15 (or LAMP-2) protein remains unexamined, but is expected to be in the area of cellular transformation given the similarity to *LMP2* of EBV. As noted above, it is evident from comparisons of exon sizes that the HHV-8 *K15* and EBV *LMP-2* genes are structurally related but, given the distinct organization of their cytoplasmic domains, they cannot be precise functional homologues. The two alleles of the K15 protein, *P* ('prototype') and *M* ('minor'), comprise 489 and 498 residues, respectively, and are extensively diverged. The non-synonymous divergence value for the *K15* alleles is 0.73 while synonymous divergence is too high to estimate. This distance between the sequences is compatible with one of the alleles (probably *M*, see below) having been introduced into HHV-8 from an Old World monkey host. No functional distinction has as yet been drawn between the *K15* alleles.

Studies have been carried out of sequence variation across the genome in HHV-8 isolates (Meng *et al*. 1999; Poole *et al*. 1999). At the left end, the extensive variation in *K1* has allowed construction of a well-developed tree (figure 6). This was originally described as having three major clades (A, B and C, with A and C most closely related) and a fourth, D, has since been added. Hayward (1999) has presented a strong case that *K1* divergence patterns correlate with the prehistoric spread of humans out of Africa and across the planet over the last 10^5 years. Clade B is seen as African, clade A as European, clade C as European and northern Asian, and clade D as southern Asian (including Australia and Oceania). At internal sites in the genome the minimal variation observed is compatible with the *K1* pattern near the left terminus but this correspondence disappears towards the right terminus and the occurrence of the *K15* alleles is essentially unlinked to the *K1* pattern. Based on the most extensive survey data available, Hayward (1999) has proposed that the *M* allele of *K15* was introduced into modern HHV-8 lineages from an unknown primate virus in a scheme involving at least two successive recombination events. Placing *K1* variation in the context of the spread and expansion of human populations over the last 10^5 years has significant implications: HHV-8 evidently did not arise by transfer into the human species in the recent past, and the relatively extended time-frame facilitates thinking about the mechanism of *K1* variation, as already outlined. However, while a period of 10^5 years constitutes the 'long term' for human society and history, it represents for both human organismal evolution and for herpesvirus evolution a short or medium time-frame, and in this context *K1* variation might perhaps be a limited term process. There are two visible factors that might be relevant to the initiation of *K1* variation: foremost is that *K1* variation might be directly consequent on the growth of human genetic diversity with global spread of our species; and second is that the virus might have been new to the human host 10^5 years ago and the extensive changes in *K1* might be an aspect of its adaptation.

5. EBV TYPES 1 AND 2

The two types of EBV, here termed EBV-1 and EBV-2, present a fascinating evolutionary case study. These two viruses show differences in their geographical and ethnic prevalences (Zimber *et al*. 1986), and in the laboratory EBV-1 strains transform lymphocytes more readily and to faster growing forms than do EBV-2 strains (Rickinson *et al*. 1987), but there is no well-defined difference in their behaviour as human pathogens. Although no complete genome sequence is available for EBV-2, a comparative overview of the two genomes can be assembled from sequences of specific genes supplemented by whole genome restriction digest data. The genomes are collinear and over most of their lengths the sequences are

extremely close, showing divergence of around 0.01 to 0.02 substitutions per site (Lees *et al.* 1993). However, there are major differences in the sequences of the EBNA-2, EBNA-3A, EBNA-3B and EBNA-3C genes (Dambaugh *et al.* 1984; Sample *et al.* 1990): non-synonymous divergences are 0.29, 0.08, 0.10 and 0.11 substitutions per site, respectively, and synonymous values 0.42, 0.13, 0.15 and 0.13 substitutions per site, respectively (McGeoch & Davison 1999*b*). The limits of the diverged regions correspond closely to the boundaries of the EBNA-2 gene and of the contiguous EBNA-3 genes. The occurrence of two types of HHV-8 *K15* gene (as discussed above) presents an apparently similar case of two well-diverged alleles of a single gene occurring in a genomic background that is generally closely conserved, although the EBV situation is intrinsically more complex with four genes involved but only two combinations of their alleles predominating.

In this section I speculate on scenarios for the genesis and propagation of EBV-1 and EBV-2. It is assumed that the two observed groupings of the EBNA-2, EBNA-3A, EBNA-3B and EBNA-3C genes possess distinct functional capabilities or advantages in particular situations, that is, that these combinations of the four-gene sets are more fit than reassorted versions and are thereby maintained. The involvement of EBNA-2 and EBNA-3C in transcriptional control and cell transformation (see Henderson *et al.* 1994) points to the likely nature of differential activities of the two virus types, although no clear understanding has emerged and the subject is not pursued here. Naturally occurring viruses with the two gene blocks reassorted have been detected (Midgley *et al.* 2000), but these could be accommodated in the rationale of functional association by supposing that they are recent recombinants that will disappear on an evolutionary time-scale.

The two EBV types are usually viewed as having arisen from recombination between a proto-EBV strain and an unknown member (or members) of the *Lymphocryptovirus* genus—the closeness of the majority of the genomic sequences would suggest that such an event took place 'recently' (say, within the last 10^5 years). Given that the two versions of the EBNA-2 gene and the two versions of the EBNA-3 gene block differ substantially in their synonymous divergence values it would be eminently reasonable to suppose that actually three distinct parental viruses might have contributed to the genome sequences of the two present day viruses. It is facile but not justified to regard EBV-1 as the proto-EBV and the EBNA-2 and EBNA-3 genes of EBV-2 as novel imports. All we can logically state about the gene compositions of the present viruses is that the histories of the EBNA-2, EBNA-3A, EBNA-3B and EBNA-3C genes differ between EBV-1 and EBV-2. Treating each set of EBNA-3 genes as a block that has a homogeneous ancestry is supported by the close similarity in divergences between the EBV-1 and EBV-2 versions of the three EBNA-3 genes as listed above, but this does remain an assumption. There is no reliable clue as to which genomic regions (either constant or variable) might possess a common line of descent that predates the putative recombination events which generated EBV-1 and EBV-2. A troubling aspect is the observed precise coincidence of the edges of EBV-1 and EBV-2 specific sequences with the four gene boundaries in the two genomic sections involved—if genesis were simply by recombination, would such precision be expected, in all of four instances? The example of extensive local variation in HHV-8 *K1* (discussed in § 4) suggests an alternative pathway by which the diverged genes of EBV-1 and EBV-2 might have arisen, by selected substitution within one virus lineage. Such a mode would avoid invocation of unknown viruses as donors and could also economically account for the distinct divergence level of the EBNA-2 gene from the EBNA-3 genes by proposing that EBNA-2 represented a longer running or more intensive instance of the divergence process. The clear and large difficulty for this mechanism is that only two alleles of the diverged EBV genes are seen, in contrast to the diaspora of HHV-8 *K1*.

Regarding the present day relationship of the two EBV types, two extreme scenarios can be considered. In the first, defined by little or no recombination between the types, EBV-1 and EBV-2 represent independent, diverging strains. In the second, recombination between the types during dual infections is taken to occur with sufficient frequency that the non-diverged major portions of the genomes will remain non-diverged, i.e. the EBNA-2 and EBNA-3 gene types are to be regarded as alleles in an otherwise homogeneous genomic setting. The latter scheme is consistent with observed sequence differences (outside the EBNA-2 and EBNA-3 genes) between EBV-1 and EBV-2 isolates being of the same magnitude as those among strains of other herpesviruses (such as herpes simplex virus) and also with studies on EBNA-1 gene sequences where variation among isolates does not correlate with assignment as EBV type 1 or 2 (Wrightham *et al.* 1995; Triantos *et al.* 1998; Habeshaw *et al.* 1999; MacKenzie *et al.* 1999). Overall, I regard this latter scheme as plausible and attractive.

A new perspective on the EBV-1–EBV-2 phenomenon has recently come from the work of Cho *et al.* (1999), who described two lymphocryptovirus types in rhesus macaques, rhesus lymphocryptovirus types 1 and 2 (RLV-1 and RLV-2) (see Wang *et al.*, this issue). The EBNA-2 gene sequences of the two RLVs are highly diverged, even more so than those of EBV-1 and EBV-2, whereas sequences of another genomic region (within the LMP-1 gene) are almost identical. Available RLV sequence data are not adequate for formal phylogenetic analysis, and it would be desirable to have data for EBNA-3 sequences and samplings from other genomic regions. Nonetheless, three points emerge with reasonable clarity. First, the genomic relationship between RLV-1 and RLV-2 appears to parallel that between EBV-1 and EBV-2. Second, comparison of EBV sequences with available RLV sequences other than the EBNA-2 gene indicates that the EBV and RLV lineages diverged in the deep past, and would be compatible with the co-speciational date of 23 Myr ago. Last, the two EBV versions of EBNA-2 genes are much closer to each other than to the RLV versions so the two pairs, EBV-1–EBV-2 and RLV-1–RLV-2, probably arose subsequent to the divergence of EBV and RLV progenitors. I regard as somewhat unattractive a mechanism of pair genesis by way of recombinational processes that in two cases, with separate host species, would require at least two previously diverged

viruses to give rise to a pair of progeny viruses having specific genomic relationships that are apparently closely equivalent in the two cases. The alternative of distinct genesis via enhanced, local variation increases in attractiveness with the RLVs' example, with parallel evolution in response to an equivalent pressure or opportunity in each host.

These speculations on the nature of the EBV-1–EBV-2 relationship should be susceptible to investigation: by evaluation of diversity levels in virus populations; by further characterization of the RLV types; by seeking more examples of the phenomenon; and, most importantly and most demanding, by attempting to study the nature of balancing selection for the two groupings of genes.

6. CONCLUSIONS

The *γ-Herpesvirinae* have in the last few years become the most active group within the *Herpesviridae* in terms of discovery of viruses, characterization of virus genomes and analysis of novel aspects of cellular interactions, in particular as relating to oncogenic transformation. This trend applies also to the phylogenetic and gene evolutionary topics described in this paper. These latter areas of study have also raised a number of complexities and difficulties, prominent among which are (i) placement of MHV-4 on the phylogenetic tree and the apparent increased rate of MHV-4 sequence change compared with other γ_2 viruses; (ii) the high rate of change in HHV-8 gene *K1* compared with the rest of its genome; and (iii) the nature of the relationship between EBV types 1 and 2. All these examples have as a component an anomaly or complication in estimated rate of sequence change, and other examples of this class mentioned in this paper include the high divergence of the HHV-8 and RRV DHFR genes compared to their HVS homologues, and the likely high divergence of primate γ_2 viruses relative to α- and β-herpesviruses. The field of γ-herpesvirus evolutionary biology is thus very open and full of interest—although an overall synthesis is certainly becoming elusive with the various phenomena and special cases now visible.

To close, I note that a more comprehensive account of the evolution of γ-herpesviruses would also need to address other topics, including (i) the deficit of the dinucleotide sequence 5′-CG in virus genomes (Honess *et al.* 1989); (ii) evolution and divergence in the HVS transforming genes (Jung *et al.* 1991); (iii) the extent to which EBV gene sequences are responsive to allele frequencies of MHC genes in human populations (De Campos-Lima *et al.* 1993; Lee *et al.* 1995; Khanna *et al.* 1997); similarly, (iv) the possible association of EBV sequence variants with neoplastic conditions (Snudden *et al.* 1995; Bhatia *et al.* 1996; Gutiérrez *et al.* 1997; MacKenzie *et al.* 1999; Habeshaw *et al.* 1999); and, lastly, (v) comparative evolutionary analysis of functional capabilities of EBV and its primate relatives (Rivailler *et al.* 1999; Ruf *et al.* 1999; Peng *et al.* 2000).

I thank Dorothy Kakoola for access to HHV-8 *K1* sequences and Andrew Davison for a critical reading of the text.

REFERENCES

Adachi, J. & Hasegawa, M. 1994 *The MOLPHY 2.2 package*. Tokyo: Institute of Statistical Mathematics.

Agius, C. T., Nagesha, H. S. & Studdert, M. J. 1992 Equine herpesvirus 5: comparisons with EHV2 (equine cytomegalovirus), cloning and mapping of a new equine herpesvirus with a novel genome structure. *Virology* **191**, 176–186.

Albrecht, J. C. 2000 Primary structure of the herpesvirus ateles genome. *J. Virol.* **74**, 1033–1037.

Albrecht, J. C. (and 10 others) 1992 Primary structure of the herpesvirus saimiri genome. *J. Virol.* **66**, 5047–5058.

Alexander, L., Denekamp, L., Knapp, A., Marcy, R., Auerbach, B. D. & Desrosiers, R. C. 2000 The primary sequence of rhesus monkey rhadinovirus isolate 26-95: sequence similarities to Kaposi's sarcoma-associated herpesvirus and rhesus monkey rhadinovirus isolate 17577. *J. Virol.* **74**, 3388–3398.

Baer, R. (and 11 others) 1984 DNA sequence and expression of the B95-8 Epstein–Barr virus genome. *Nature* **310**, 207–211.

Bhatia, K., Raj, A., Gutiérrez, M. I., Judde, J. G., Spangler, G., Venkatesh, H. & Magrath, I. T. 1996 Variation in the sequence of Epstein–Barr virus nuclear antigen 1 in normal peripheral blood lymphocytes and in Burkitt's lymphomas. *Oncogene* **13**, 177–181.

Broll, H., Finsterbusch, T., Buhk, H. J. & Goltz, M. 1999 Genetic analysis of the bovine herpesvirus type 4 gene locus for the putative terminase. *Virus Genes* **19**, 243–250.

Cameron, K. R., Stamminger, T., Craxton, M., Bodemer, W., Honess, R. W. & Fleckenstein, B. 1987 The 160,000-M_r virion protein encoded at the right end of the herpesvirus saimiri genome is homologous to the 140,000-M_r membrane antigen encoded at the left end of the Epstein–Barr virus genome. *J. Virol.* **61**, 2063–2070.

Cho, Y. G., Gordadze, A. V., Ling, P. D. & Wang, F. 1999 Evolution of two types of rhesus-lymphocryptovirus similar to type 1 and type 2 Epstein–Barr virus. *J. Virol.* **73**, 9206–9212.

Cook, P. M., Whitby, D., Calabro, M. L., Luppi, M., Kakoola, D. N., Hjalgrim, H., Ariyoshi, K., Ensoli, B., Davison, A. J. & Schulz, T. F. 1999 Variability and evolution of Kaposi's sarcoma-associated herpesvirus in Europe and Africa. *AIDS* **13**, 1165–1176.

Dambaugh, T., Hennessy, K., Chammankit, L. & Kieff, E. 1984 U2 region of Epstein–Barr virus DNA may encode Epstein–Barr nuclear antigen 2. *Proc. Natl Acad. Sci. USA* **81**, 7632–7636.

Davison, A. J. & Scott, J. E. 1986 The complete DNA sequence of varicella-zoster virus. *J. Gen. Virol.* **67**, 1759–1816.

De Campos-Lima, P. O., Gavioli, P., Zhang, Q. J., Wallace, L. E., Dolcetti, R., Rowe, M., Rickinson, A. B. & Masucci, M. G. 1993 HLA-A11 epitope loss isolates of Epstein–Barr virus from a highly $A11^+$ population. *Science* **260**, 98–100.

Ehlers, B., Borchers, K., Grund, C., Frölich, K., Ludwig, H. & Buhk, H. J. 1999 Detection of new DNA polymerase genes of known and potentially novel herpesviruses by PCR with degenerate and deoxyinosine-substituted primers. *Virus Genes* **18**, 211–220.

Ensser, A., Pflanz, R. & Fleckenstein, B. 1997 Primary structure of the alcelaphine herpesvirus 1 genome. *J. Virol.* **71**, 6517–6525.

Goltz, M., Broll, H., Mankertz, A., Weigelt, W., Ludwig, H., Buhk, H. J. & Borchers, K. 1994 Glycoprotein B of bovine herpesvirus type 4: its phylogenetic relationship to gB equivalents of the herpesviruses. *Virus Genes* **9**, 53–59.

Greensill, J., Sheldon, J. A., Renwick, N. W., Beer, B. E., Norley, S., Goudsmit, J. & Schulz, T. F. 2000 Two distinct gamma-2 herpesviruses in African green monkeys: a second

gamma-2 herpesvirus lineage among Old World primates. *J. Virol.* **74**, 1572–1577.

Gutiérrez, M. I. (and 10 others) 1997 Sequence variations in EBNA-1 may dictate restriction of tissue distribution of Epstein–Barr virus in normal and tumour cells. *J. Gen. Virol.* **78**, 1663–1670.

Habeshaw, G., Yao, Q. Y., Bell, A. I., Morton, D. & Rickinson, A. B. 1999 Epstein–Barr virus nuclear antigen 1 sequences in endemic and sporadic Burkitt's lymphoma reflect virus strains prevalent in different geographic areas. *J. Virol.* **73**, 965–975.

Hayward, G. S. 1999 KSHV strains: the origins and global spread of the virus. *Semin. Cancer Biol.* **9**, 187–199.

Henderson, S. A., Huen, D. & Rowe, M. 1994 Epstein–Barr virus transforming proteins. *Semin. Virol.* **5**, 391–399.

Holloway, S. A., Lindquester, G. J., Studdert, M. J. & Drummer, H. E. 1999 Identification, sequence analysis and characterization of equine herpesvirus 5 glycoprotein B. *Arch. Virol.* **144**, 287–307.

Honess, R. W., Gompels, U. A., Barrell, B. G., Craxton, M., Cameron, K. R., Staden, R., Chang, Y. N. & Hayward, G. S. 1989 Deviations from expected frequencies of CpG dinucleotides in herpesvirus DNAs may be diagnostic of differences in the states of their latent genomes. *J. Gen. Virol.* **70**, 837–855.

Hudson, R. R. 1990 Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* **7**, 1–44.

Ishido, S., Wang, C., Lee, B. S., Cohen, G. B. & Jung, J. U. 2000*a* Downregulation of major histocompatibility complex class I molecules by Kaposi's sarcoma-associated herpesvirus K3 and K5 proteins. *J. Virol.* **74**, 5300–5309.

Ishido, S., Choi, J.-K., Lee, B. S., Wang, C., DeMaria, M., Johnson, R. P., Cohen, G. B. & Jung, J. U. 2000*b* Inhibition of natural killer cell-mediated cytotoxicity by Kaposi's sarcoma-associated herpesvirus K5 protein. *Immunity* **13**, 365–374.

Jones, D. T., Taylor, W. R. & Thornton, J. M. 1992 The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282.

Jung, J. U., Trimble, J. J., King, N. W., Biesinger, B., Fleckenstein, B. W. & Desrosiers, R. C. 1991 Identification of transforming genes of subgroup A and C strains of herpesvirus saimiri. *Proc. Natl Acad. Sci. USA* **88**, 7051–7055.

Kaplan, N., Darden, T. & Hudson, R. R. 1988 The coalescent process in models with selection. *Genetics* **120**, 819–829.

Kaplan, N. L., Hudson, R. R. & Langley, C. H. 1989 The 'hitchhiking effect' revisited. *Genetics* **123**, 887–899.

Khanna, R., Slade, R. W., Poulsen, L., Moss, D. J., Burrows, S. R., Nicholls, J. & Burrows, J. M. 1997 Evolutionary dynamics of genetic variation in Epstein–Barr virus isolates of diverse geographical origins: evidence for immune pressure-independent genetic drift. *J. Virol.* **71**, 8340–8346.

Koseki, T., Inohara, N., Chen, S., Carrio, R., Merino, J., Hottiger, M. O., Nable, G. J. & Nunez, G. 1999 CIPER, a novel NFκB-activating protein containing a caspase recruitment domain with homology to herpesvirus-2 protein E10. *J. Biol. Chem.* **274**, 9955–9961.

Kumar, S. & Hedges, S. B. 1998 A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920.

Lacoste, V., Mauclère, P., Dubreuil, G., Lewis, J., Georges-Courbot, M.-C. & Gessain, A. 2000 KSHV-like herpesviruses in chimps and gorillas. *Nature* **407**, 151–152.

Lee, S. P., Morgan, S., Skinner, J., Thomas, W. A., Jones, S. R., Sutton, J., Khanna, R., Whittle, H. C. & Rickinson, A. B. 1995 Epstein–Barr virus isolates with the major HLA B35.01-restricted cytotoxic T-lymphocyte epitope are prevalent in a highly B35.01-positive African population. *Eur. J. Immunol.* **25**, 102–110.

Lee, H., Veazey, R., Williams, K., Li, M., Guo, J., Neipel, F., Fleckenstein, B., Lackner, A., Desrosiers, R. C. & Jung, J. U. 1998 Deregulation of cell growth by the K1 gene of Kaposi's sarcoma associated herpesvirus. *Nature Med.* **4**, 435–440.

Lees, J. F., Arrand, J. E., De V. Pepper, S., Stewart, J. P., Mackett, M. & Arrand, J. R. 1993 The Epstein–Barr virus candidate vaccine antigen gp340/220 is highly conserved between virus types A and B. *Virology* **195**, 578–586.

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

McGeoch, D. J. & Davison, A. J. 1999*a* The descent of human herpesvirus 8. *Semin. Cancer Biol.* **9**, 201–209.

McGeoch, D. J. & Davison, A. J. 1999*b* The molecular evolutionary history of the herpesviruses. In *Origins and evolution of viruses* (ed. E. Domingo, R. Webster & J. Holland), pp. 441–465. London: Academic Press.

McGeoch, D. J., Cook, S., Dolan, A., Jamieson, F. E. & Telford, E. A. R. 1995 Molecular phylogeny and evolutionary timescale for the family of mammalian herpesviruses. *J. Mol. Biol.* **247**, 443–458.

McGeoch, D. J., Dolan, A. & Ralph, A. C. 2000 Toward a comprehensive phylogeny for mammalian and avian herpesviruses. *J. Virol.* **74**, 10 401–10 406.

MacKenzie, J., Gray, D., Pinto-Paes, R., Barrezueta, L. F. M., Armstrong, A., Alexander, F. A., McGeoch, D. J. & Jarrett, R. F. 1999 Analysis of Epstein–Barr virus (EBV) nuclear antigen 1 subtypes in EBV-associated lymphomas from Brazil and the United Kingdom. *J. Gen. Virol.* **80**, 2741–2745.

Meng, Y. X. (and 11 others) 1999 Individuals from North America, Australasia and Africa are infected with four different genotypes of human herpesvirus 8. *Virology* **261**, 106–119.

Midgley, R. S., Blake, N. W., Yao, Q. Y., Croom-Carter, D., Cheung, S. F., Chan, A. T. C., Johnson, P. J., Huang, D., Rickinson, A. B. & Lee, S. P. 2000 Novel intertypic recombinants of Epstein–Barr virus in the Chinese population. *J. Virol.* **74**, 1544–1548.

Minson, A. C., Davison, A. J., Desrosiers, R. C., Fleckenstein, B., McGeoch, D. J., Pellett, P. E., Roizman, B. & Studdert, D. M. J. 2000 Herpesviridae. In *Virus taxonomy* (ed. M. H. V. van Regenmortel and 10 others), pp. 203–225. New York and San Diego, CA: Academic Press.

Neipel, F. & Fleckenstein, B. 1999 The role of HHV-8 in Kaposi's sarcoma. *Semin. Cancer Biol.* **9**, 151–164.

Neipel, F., Albrecht, J. C. & Fleckenstein, B. 1997 Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus human herpesvirus 8: determinants of its pathogenicity? *J. Virol.* **71**, 4187–4192.

Nicholas, J. (and 11 others) 1998 Novel organizational features, captured cellular genes, and strain variability within the genome of KSHV/HHV8. *Monogr. Natl Cancer Inst.* **23**, 79–88.

Parker, B. D., Bankier, A., Satchwell, S., Barrell, B. & Farrell, P. J. 1990 Sequence and transcription of Raji Epstein–Barr virus DNA spanning the B95-8 deletion. *Virology* **179**, 339–346.

Peng, R. S., Gordadze, A. V., Pananá, E. M. F., Wang, F., Zong, J., Hayward, G. S., Tan, J. & Ling, P. D. 2000 Sequence and functional analysis of EBNA-LP and EBNA2 proteins from non-human primate lymphocryptoviruses. *J. Virol.* **74**, 379–389.

Poole, L. J., Zong, J. C., Ciufo, D. M., Alcendor, D. J., Cannon, J. S., Ambinder, R., Orenstein, J. M., Reitz, M. S. & Hayward, G. S. 1999 Comparison of genetic variability at multiple loci across the genomes of the major subtypes of Kaposi's sarcoma associated herpesvirus reveals evidence for recombination and for two distinct types of open reading frame K15 alleles at the right-hand end. *J. Virol.* **73**, 6646–6660.

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.

Rickinson, A. B., Young, L. S. & Rowe, M. 1987 Influence of Epstein–Barr virus nuclear antigen EBNA2 on the growth phenotype of virus transformed B cells. *J. Virol.* **61**, 1310–1317.

Rivailler, P., Quink, C. & Wang, F. 1999 Strong selective pressure for evolution of an Epstein–Barr virus LMP2B homologue in the rhesus lymphocryptovirus. *J. Virol.* **73**, 8867–8872.

Rovnak, J., Quackenbush, S. L., Reyes, R. A., Baines, J. D., Parrish, C. R. & Casey, J. W. 1998 Detection of a novel bovine lymphotropic herpesvirus. *J. Virol.* **72**, 4237–4242.

Ruf, I. K., Moghaddam, A., Wang, F. & Sample, J. 1999 Mechanisms that regulate Epstein–Barr virus EBNA-1 gene transcription during restricted latency are conserved among lymphocryptoviruses of Old World primates. *J. Virol.* **73**, 1980–1989.

Russo, J. J. (and 10 others) 1996 Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). *Proc. Natl Acad. Sci. USA* **93**, 14 862–14 867.

Sample, J., Young, L., Martin, B., Chatman, T., Kieff, E., Rickinson, A. & Kieff, E. 1990 Epstein–Barr virus types 1 and 2 differ in their EBNA-3A, EBNA-3B and EBNA-3C genes. *J. Virol.* **64**, 4084–4092.

Schultz, E. R., Rankin, J. G. W., Blanc, M. P., Raden, B. W., Tsai, C. C. & Rose, T. M. 2000 Characterization of two divergent lineages of macaque rhadinoviruses related to Kaposi's sarcoma-associated herpesvirus. *J. Virol.* **74**, 4919–4928.

Searles, R. P., Bergquam, E. P., Axthelm, M. K. & Wong, S. W. 1999 Sequence and genomic analysis of a rhesus macaque rhadinovirus with similarity to Kaposi's sarcoma-associated herpesvirus/human herpesvirus 8. *J. Virol.* **73**, 3040–3053.

Snudden, D. K., Smith, P. R., Lai, D., Ng, M. H. & Griffin, B. E. 1995 Alterations in the structure of the EBV nuclear antigen, EBNA1, in epithelial cell tumours. *Oncogene* **10**, 1545–1552.

Telford, E. A. R., Watson, M. S., Aird, H. C., Perry, J. & Davison, A. J. 1995 The DNA sequence of equine herpesvirus 2. *J. Mol. Biol.* **249**, 520–528.

Thome, M., Martinon, F., Hofmann, K., Rubio, V., Steiner, V., Schneider, P., Mattmann, C. & Tschopp, J. 1999 Equine herpesvirus 2 E10 gene product, but not its cellular homologue, activates NF-κB transcription factor and c-Jun N-terminal kinase. *J. Biol. Chem.* **274**, 9962–9968.

Triantos, D., Boulter, A. W., Leao, J. C., Di Alberti, L., Porter, S. R., Scully, C. M., Birnbaum, W., Johnson, N. W. & Teo, C. G. 1998 Diversity of naturally occurring Epstein–Barr virus revealed by nucleotide sequence polymorphism in hypervariable domains in the *Bam*HI K and N subgenomic regions. *J. Gen. Virol.* **79**, 2809–2817.

Ulrich, S., Goltz, M. & Ehlers, B. 1999 Characterization of the DNA polymerase loci of the novel porcine lymphotropic herpesvirus 1 and 2 in domestic and feral pigs. *J. Gen. Virol.* **80**, 3199–3205.

Van Santen, V. L. 1991 Characterization of the bovine herpesvirus 4 major immediate-early transcript. *J. Virol.* **65**, 5211–5224.

Vanderplasschen, A. (and 10 others) 2000 A multipotential β-1,6-*N*-acetylglucosaminyltransferase is encoded by bovine herpesvirus type 4. *Proc. Natl Acad. Sci. USA* **97**, 5756–5761.

Virgin, H. W., Latreille, P., Wamsley, P., Hallsworth, K., Weck, K. E., Dal Canto, A. J. & Speck, S. H. 1997 Complete sequence and genomic analysis of murine gammaherpesvirus 68. *J. Virol.* **71**, 5894–5904.

Wrightham, M. N., Stewart, J. P., Janjua, N. J., Pepper, S. D. V., Sample, C., Rooney, C. M. & Arrand, J. R. 1995 Antigenic and sequence variation in the C-terminal unique domain of the Epstein–Barr virus nuclear antigen EBNA-1. *Virology* **208**, 521–530.

Yang, Z. 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556.

Yates, J. L., Camiolo, S. M., Ali, S. & Ying, A. 1996 Comparison of the EBNA1 proteins of Epstein–Barr virus and herpesvirus papio in sequence and function. *Virology* **222**, 1–13.

Zimber, U. (and 10 others) 1986 Geographical prevalence of two Epstein–Barr virus types. *Virology* **154**, 56–66.

Zong, J. C. (and 16 others) 1999 High-level variability in the ORF-K1 membrane protein gene at the left end of the Kaposi's sarcoma-associated herpesvirus genome defines four major virus subtypes and multiple variants or clades in different human populations. *J. Virol.* **73**, 4156–4170.